

22082
PCTORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁶ : G06T 15/70	A1	(11) Numéro de publication internationale: WO 97/46974 (43) Date de publication internationale: 11 décembre 1997 (11.12.97)
<p>(21) Numéro de la demande internationale: PCT/FR97/00981</p> <p>(22) Date de dépôt international: 3 juin 1997 (03.06.97)</p> <p>(30) Données relatives à la priorité: 96/06813 3 juin 1996 (03.06.96) FR</p> <p>(71)(72) Déposants et inventeurs: PRONIER, Jean-Luc [FR/FR]; 108, rue de la République, F-95100 Argenteuil (FR). ALFONSI, Philippe [FR/FR]; 9, rue Falguière, F-75015 Paris (FR).</p> <p>(74) Représentant commun: PRONIER, Jean-Luc; 108, rue de la République, F-95100 Argenteuil (FR).</p>		<p>(81) Etats désignés: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, brevet ARIPO (GH, KE, LS, MW, SD, SZ, UG), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Publiée <i>Avec rapport de recherche internationale.</i></p>
<p>(54) Title: DEVICE AND METHOD FOR TRANSMITTING ANIMATED AND SOUND IMAGES</p> <p>(54) Titre: DISPOSITIF ET PROCEDE DE TRANSMISSION D'IMAGES ANIMEES ET SONORISEES</p> <p>(57) Abstract</p> <p>The invention discloses a method for forming animated and sound images representing the face of a speaker consisting in effecting the following steps: initially: receiving an image of the said speaker's face; modeling a face said to be "modelled" representing the said face and adapted to be animated; and memorising the said modelled face; then after receiving a message likely to be spoken by the said speaker, analysing this message to determine a series of facial expressions to match the speaking out of the said message, and: transmitting a voice speaking out the said message, and animating the modelled face corresponding all the time to the delivery of the message.</p> <p>(57) Abrégé</p> <p>Le procédé de formation d'images animées et sonorisées représentatives du visage d'un interlocuteur consiste à effectuer successivement les étapes suivantes: initialement: recevoir une image d'un visage dudit interlocuteur; modéliser un visage dit "modéliser" représentatif dudit visage et adapté à être animé; et mémoriser ledit visage modéliser; puis, après avoir reçu un message susceptible d'être prononcé par ledit interlocuteur, analyser ce message pour déterminer une succession d'expressions de visage qui correspond à la prononciation dudit message, et: émettre une voix prononçant ledit message, et animer le visage modélisé correspondant à chaque instant à la prononciation du message.</p>		

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	IS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brsil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakhstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

DISPOSITIF ET PROCÉDE DE TRANSMISSION D'IMAGES
ANIMEES ET SONORISEES.

5 La présente invention concerne un dispositif et
un procédé de transmission d'images animées et sonorisées
représentatives d'au moins un visage d'une personne.

 Dans la suite, le mot « personne » est
équivalent au mot « interlocuteur » et désigne la personne
10 dont une image du visage est transmise.

 L'invention s'applique aussi bien à la
formation d'images à distance, par exemple pour des
téléspectateurs, des auditeurs de stations radiophoniques
ou des utilisateurs de télévisiophones, qu'à la formation
15 d'images localement pour l'insertion dans un jeu vidéo.

 Les dispositifs de transmission d'images
animées actuellement connus nécessitent l'utilisation d'un
capteur photosensible qui fournit un signal électrique
représentant des luminosités perçues. La quantité
20 d'information représentative de l'image est alors très
élevée et, bien que des technologies de compression
d'images existent, la transmission des images nécessite un
support de transmission capable de transférer de grandes
quantités d'informations par seconde.

25 De plus, l'utilisation du capteur photosensible
implique de maîtriser les conditions de prise de vue,
telles que l'éclairage, le diaphragme, la longueur focale
et le réglage de la mise au point. Enfin, la capture et la
transmission d'images animées est alors d'un coût élevé à
30 cause des quantités d'informations à transmettre.

On connaît le document « Systems & Computers in Japan », volume 22, no. 5, 1991, New York U.S. pages 50 à 59, qui présente des études de synthèse de mouvements faciaux pour une interface homme-machine intelligente. Ces études tendent à fournir une interface possédant une image faciale et une voix synchronisées. La face présente des expressions dont la synthèse utilise des données de texte et des émotions. Cependant, cette interface ne permet pas réaliser une animation en temps réel en partant de la voix d'un interlocuteur puisque ce sont des données d'un fichier informatique de texte qui servent d'une part à une synthèse de voix et, d'autre part, à animer la bouche du personnage de l'interface. En outre, ce document ne suggère aucune correspondance entre la personne qui a produit le texte et la personne dont la face est représentée par l'interface. Enfin, ce document ne suggère aucune transmission à distance de la face ainsi synthétisée et animée par des données de texte.

On connaît aussi le document GB-2 250 405 A qui présente un dispositif d'analyse de voix et de synthèse d'image. Ce dispositif analyse des séquences vocale pour produire des séquences de mots de code identifiant la lettre prononcée, puis la probabilité qu'une voix corresponde à une forme de bouche. Ce document ne suggère aucune correspondance entre la personne qui a produit la voix analysée et l'image de synthèse fournissant une face parlante (« talking face »). En outre, ce document ne suggère aucune transmission à distance de la face ainsi synthétisée et animée par des données de texte.

On connaît encore le document IEEE Virtual reality annual international symposium, 18-22 septembre

1993, New York NY US pages 486-491. Ce document décrit des recherches dont le but est de donner une face à une terminal informatique. Cette face provenant d'un mannequin est animée par une voix naturelle ou un texte. Ce document
5 ne suggère aucune correspondance entre la personne qui a produit le texte ou la voix et la personne dont la face est représentée par l'interface. Enfin, ce document ne suggère aucune transmission à distance de la face ainsi synthétisée et animée par des données de texte.

10 On connaît, enfin, le document INTERCHI '93 Conference proceedings, 24-29 Avril 1993, Amsterdam, page 187-193. Ce document présente des recherches concernant des écrans faciaux et des modalités de conversation. Il analyse la qualité de perception de messages selon qu'ils sont
15 appuyés par une face animée ou non. Cependant, cette interface ne permet pas réaliser une animation en temps réel en partant de la voix d'un interlocuteur puisque la voix n'est pas analysée. En outre, ce document ne suggère aucune correspondance entre la personne qui a produit le
20 texte et la personne dont la face est représentée par l'interface. Enfin, ce document ne suggère aucune transmission à distance de la face ainsi synthétisée et animée par des données de texte.

La présente invention entend remédier à ces
25 inconvénients en proposant de modéliser une image d'un visage d'un interlocuteur, pour former un visage modélisé qui peut être animé, d'analyser un message de cet interlocuteur pour déterminer une expression faciale correspondant à une voix prononçant ce message puis
30 d'animer le modèle d'image de visage pour lui donner ladite expression faciale tout en émettant la voix.

Grâce à ces dispositions :

- l'animation peut être effectuée en temps réel puisqu'elle utilise des signaux correspondants à une voix, et donc une quantité d'information suffisamment faible pour permettre un traitement rapide, si ce n'est instantané ;
- le visage et la voix correspondent au même interlocuteur ; et
- l'image du visage de l'interlocuteur considéré est animée à distance, par le signal représentatif de message de cet interlocuteur.

Ainsi, la capture des mouvements et expressions du visage de l'interlocuteur est effectuée en captant non pas des rayons lumineux réfléchis par le visage mais un message susceptible d'être prononcé par ledit visage, et plus particulièrement par un message vocal prononcé par la bouche, elle même représentative de l'expression du visage. Le coût du dispositif de l'invention est ainsi limité à celui de réalisation d'une image fixe, par exemple avec un appareil photographique ou par ordinateur, et à celui de capture d'un message, par exemple par un appareil téléphonique ou par un ordinateur.

Selon un premier aspect, la présente invention vise un dispositif de transmission d'images animées et sonorisées représentatives d'un interlocuteur, caractérisé en ce qu'il comporte :

- un moyen de mémorisation d'une image d'un visage dit « modélisé » représentatif du visage dudit interlocuteur ;

- un moyen de communication à distance de signaux représentatifs d'un message susceptible d'être prononcé par ledit interlocuteur ;
- un moyen d'analyse desdits signaux adapté à déterminer
5 une succession d'expressions de visage correspondant à la prononciation dudit message par une voix, selon des règles prédéterminées prenant en compte lesdits signaux ;
- un moyen d'animation adapté à animer l'image du visage
10 modélisé de l'interlocuteur pour que ledit visage présente successivement chaque expression de ladite succession d'expressions de visage correspondant à la prononciation dudit message par ladite voix, et
- un moyen d'affichage d'image et d'émission de sons
15 vocaux, adapté à émettre ladite voix et à afficher simultanément ledit visage modélisé présentant l'expression correspondant à chaque instant à la prononciation dudit message par ladite voix.

Selon un deuxième aspect, la présente invention
20 vise un dispositif de transmission d'images animées et sonorisées représentatives du visage d'un interlocuteur, caractérisé en ce qu'il comporte :

- un moyen de mémorisation d'une image d'un visage dit
« modélisé » représentatif du visage dudit
25 interlocuteur ;
- un moyen de réception de signaux représentatifs d'un message susceptible d'être prononcé par ledit interlocuteur ;
- un moyen d'analyse desdits signaux adapté à déterminer
30 une succession d'expressions de visage correspondant à la prononciation dudit message par une voix, selon des

5 règles prédéterminées prenant en compte lesdits signaux ;

- un moyen d'animation adapté à animer l'image du visage modélisé de l'interlocuteur pour que ledit visage présente successivement chaque expression de ladite succession d'expressions de visage correspondant à la prononciation dudit message par ladite voix, et
- un moyen de transmission de signaux simultanément représentatifs de sons vocaux, correspondant à ladite

10 voix et d'images correspondant audit visage modélisé présentant l'expression correspondant à chaque instant à la prononciation dudit message par ladite voix.

Grâce à chacune de ces dispositions, le support de transmission d'images animées est un support de

15 transmission de voix ou de texte et il n'est donc pas nécessaire qu'il soit capable de transmettre plus que les fréquences vocales. La complexité et le coût de cette transmission et de la formation d'images animées sont donc très limités.

20 Selon des caractéristiques particulières, dans le dispositif de transmission tel que succinctement exposé ci-dessus, le moyen de transmission à distance est adapté à transmettre des signaux représentatifs d'un message vocal prononcé par ledit interlocuteur.

25 Grâce à ces dispositions, la capture d'un message vocal prononcé par l'interlocuteur, par exemple par utilisation d'un microphone, suffit pour provoquer l'animation, à distance, d'un visage représentatif de cet interlocuteur.

30 Selon des caractéristiques particulières, dans le dispositif de transmission tel que succinctement exposé

ci-dessus, le moyen de transmission à distance est adapté à transmettre des signaux représentatifs d'un texte susceptible d'être prononcé par l'interlocuteur.

Grâce à ces dispositions, la capture d'un texte susceptible d'être prononcé par l'interlocuteur, par exemple par utilisation d'un clavier, suffit pour provoquer l'animation, à distance, d'un visage représentatif de cet interlocuteur.

Selon des caractéristiques particulières, le dispositif de transmission tel que succinctement exposé ci-dessus, comporte, en outre, un moyen de modélisation d'image adapté à fournir une image destinée à être animée, en fonction d'une image prise par un capteur photosensible et en ce que le moyen de mémorisation mémorise ladite image.

Grâce à ces dispositions, la capture d'une image électronique, par exemple par une caméra électronique ou un scanner suffit à la mémorisation de l'image destinée à être modélisée. Ainsi, la transmission d'une image fixe destinée à être animée est réalisée par transmission d'une photographie ou d'une image vidéo, et le coût de capture et de transmission est très limité.

Selon des caractéristiques particulières, le dispositif de transmission tel que succinctement exposé ci-dessus, comporte, en outre, un récepteur téléphonique relié à une ligne téléphonique.

Grâce à ces dispositions, les signaux représentatifs du message et/ou des signaux représentatifs de sons vocaux et d'images peuvent être véhiculés à grande distance, par exemple sur un réseau téléphonique ou

informatique, qu'il soit commuté ou numérique. En outre, un service de télévisiophonie peut ainsi être mis en place.

Selon des caractéristiques particulières, le dispositif de transmission tel que succinctement exposé ci-dessus, comporte, en outre, une antenne hertzienne.

Grâce à ces dispositions :

- des signaux radiophoniques peuvent permettre l'animation à distance de visages représentatifs des visages des interlocuteurs de la station radiophonique, et/ou
- des émissions télévisuelles dans lesquelles peuvent intervenir des images modélisées des téléspectateurs peuvent être diffusées.

L'invention vise aussi une console de jeu, un ordinateur, un banc de montage audiovisuel, un télévisiophone caractérisés en ce qu'ils comportent un dispositif de transmission tel que succinctement présenté ci-dessus.

Selon un troisième aspect, la présente invention vise un procédé de formation d'images animées et sonorisées représentatives du visage d'un interlocuteur, caractérisé en ce qu'il consiste à effectuer successivement les étapes suivantes :

- dans un premier temps :
 - * une opération de réception d'une image d'un visage dudit interlocuteur ;
 - * une opération de modélisation d'un visage dit « modélisé » représentatif dudit visage et adapté à être animé ; et
 - * une opération de mémorisation dudit visage modélisé ;

- puis, dans un deuxième temps et après une opération de réception de signaux représentatifs d'un message susceptible d'être prononcé par ledit interlocuteur,

5 * une opération d'analyse de ces signaux pour déterminer une succession d'expressions de visage qui correspond à la prononciation dudit message par une voix ; et, simultanément :

. une opération d'émission sonore de ladite voix prononçant ledit message, et

10 . une opération d'animation dudit visage modélisé pour qu'il présente successivement chaque expression de ladite succession d'expressions de visage correspondant à chaque instant à la prononciation dudit message par ladite voix.

15 Grâce à chacune de ces dispositions, l'étape réalisée dans un premier temps peut être faite préliminairement à la réception de la voix, éventuellement assez lentement. La puissance de calcul nécessaire est ainsi moins élevée que pour la transmission d'une image
20 animée captée par un capteur photosensible.

Ainsi peut être effectuée la transmission dans une émission de télévision d'une image animée d'un téléspectateur qui participe à l'émission.

Les avantages du procédé visé selon le
25 troisième aspect de l'invention comportent, en outre, ceux des dispositifs tel que succinctement exposés ci-dessus.

L'invention vise aussi une console de jeu, un ordinateur, un banc de montage audiovisuel et un télévisiophone, caractérisés en ce qu'ils mettent en oeuvre
30 un procédé de transmission d'images tel que succinctement présenté ci-dessus.

D'autres avantages, buts et caractéristiques de la présente invention ressortiront de la description qui va suivre, faite en regard des dessins annexés, dans lesquels :

5 - la figure 1 représente un premier mode de réalisation de la présente invention, mis en oeuvre dans une production télévisuelle ;

 - la figure 2 représente un deuxième mode de réalisation de la présente invention, mis en oeuvre dans
10 une station radiophonique ;

 - la figure 3 représente un troisième mode de réalisation de la présente invention, mis en oeuvre dans un réseau informatique ;

 - la figure 4 représente un quatrième mode de réalisation de la présente invention, mis en oeuvre dans
15 un réseau informatique ;

 - la figure 5 représente un schéma bloc de dispositifs d'émission d'images mis en oeuvre dans le premier mode de réalisation de la présente invention ;

20 - la figure 6 représente les opérations successives réalisées par les dispositifs illustrés en figure 5 ;

 - la figure 7 représente une implantation matérielle d'un dispositif présenté en figure 5 ;

25 - la figure 8 représente un visage d'enfant photographié ;

 - la figure 9 représente la réalisation d'une image modélisée du visage présenté en figure 8 ;

 - la figure 10 représente le spectre d'un
30 signal sonore représentatif d'une voix prononçant un phonème particulier ;

- la figure 11 représente l'expression correspond au signal sonore présenté en figure 10 ;

- la figure 12 représente un modèle animé par la voix dont le spectre est représenté en figure 10, destiné à être mis en correspondance avec le visage représenté sur la photographie présentée en figure 8 pour fournir l'image de l'interlocuteur.

En figure 1, le premier mode de réalisation est mis en oeuvre pour une production d'oeuvres télévisuelles dans laquelle un téléspectateur peut intervenir, son image animée par sa voix étant diffusée en direct dans le cadre de l'émission. Ce téléspectateur, aussi appelé « interlocuteur » dans la suite de la description, utilise un téléphone comportant un combiné 101 muni d'un microphone 102 et un clavier 103. Ce téléphone est relié à un réseau téléphonique 105 par une prise téléphonique 104.

Dans le studio d'enregistrement de l'émission, une prise téléphonique 106 permet de recevoir les signaux émis par le téléphone du téléspectateur considéré. Un moyen d'analyse de signaux sonores 107 analyse les signaux reçus sur la prise téléphonique 106, et fournit des informations représentatives d'expressions du visage de l'interlocuteur correspondant à la prononciation du message vocal porté par lesdits signaux sonores. Le fonctionnement du moyen d'analyse 107 est détaillé en regard de la figure 5.

Un moyen de capture d'image 108, ici constitué d'un scanner de bureau, a préliminairement permis de fournir une image électronique numérique d'au moins une photographie de visage que l'interlocuteur a envoyé, par la poste ou par télécopie.

De manière manuelle ou par un traitement d'image automatique, l'image électronique numérique a été associée à un visage dit « modélisé » constitué d'un modèle en trois dimensions qui représente le visage de l'interlocuteur et possède des points caractéristiques dont les déplacements permettent de faire présenter à ce visage des expressions faciales. Ces points, sont, par exemple, les commissures des lèvres, les points les plus haut et les plus bas des lèvres, les coins des yeux.

Un moyen de mémorisation de modèle 109, ici constitué d'une mémoire de masse, de type disque dur, d'un terminal informatique, conserve les données nécessaires à la reconstruction du visage dit modélisé.

Un moyen d'animation de modèle 110 anime le visage modélisé qui est conservé dans le moyen de mémorisation de modèle 109 pour lui donner les expressions de visage fournies par le moyen d'analyse de signaux sonores 107. A cet effet, il déplace les points caractéristiques du visage modélisé selon des règles de déplacement connues et les autres points du visage modélisé sont déplacés selon des règles de déformation connues. Pour la compréhension de ces déplacements, on pourra se reporter à la lecture des actes et proceedings de la conférence IMAGINA de l'année 1997 et, en particulier à ses pages 246 à 257 qui exposent les travaux effectués à l'Institut National de l'Audiovisuel.

L'image animée du visage modélisé représentatif du visage de l'interlocuteur et le signal sonore représentatif de la voix de l'interlocuteur sont diffusés simultanément par un émetteur hertzien 111, comportant une antenne hertzienne et émettant à destination d'une

multitude de récepteurs de télévision reliés à des antennes hertziennes, en faisant correspondre, à tout moment, l'expression du visage modélisé à la prononciation de la voix de l'interlocuteur.

5 Chaque récepteur de télévision 112 est muni d'un écran de visualisation 114 et d'un haut-parleur 113. Il diffuse, selon des techniques télévisuelles connues :

- par l'intermédiaire de haut-parleur 113, la voix de l'interlocuteur, éventuellement combinée à d'autres
- 10 signaux sonores caractéristiques de l'émission, et
- par l'intermédiaire de l'écran de visualisation 114, le visage animé de l'interlocuteur, éventuellement complété par un décor, des personnages et une mise en
- scène caractéristiques de l'émission.

15 Le schéma fonctionnel et le fonctionnement du premier mode de réalisation illustré en figure 1 sont détaillés en regard de la figure 5.

 Selon une variante, le moyen de capture d'image 108 est un modem servant à la réception d'un fichiers image

20 émis par un ordinateur ou un télécopieur.

 En figure 2, un deuxième mode de réalisation de la présente invention est mis en oeuvre dans une station radiophonique. Chaque animateur ou invité de la station considérée, aussi appelé « interlocuteur » dans la suite de

25 la description, est entouré de trois microphones 120, 121 et 122 reliés à un moyen de traitement de sons 133, et dans le champ optique d'une caméra 123 associée à un moyen de traitement d'images 124.

 Le moyen de traitement de sons 133, le moyen de

30 traitement d'images 124 et une console numérique 125 sont reliés à un moyen d'insertion de données numériques 134

lui-même relié à un émetteur hertzien 126 comportant une antenne hertzienne.

Le microphone 120 est placé à droite de l'animateur à hauteur de sa bouche. Le microphone 122 est
5 placé à gauche de l'animateur à hauteur de sa bouche. Le microphone 121 est placé au dessus de la tête de l'animateur dans l'axe médian des deux autres microphones.

Le moyen de traitement de sons 133 analyse les ratios des intensités sonores représentées par les signaux
10 émis par chacun des trois microphones et fournit une information représentative de la position de la tête de l'animateur. En effet, plus la tête de l'animateur est tournée vers l'un des microphones, plus l'intensité sonore reçue par ce microphone est élevée et plus l'intensité
15 sonore reçue par les autres microphones est faible.

La caméra 123 comporte un capteur électronique qui fournit un signal représentatif de l'image du visage de l'animateur, selon des techniques connues dans le domaine des caméras vidéos. Le moyen de traitement d'image 124
20 analyse le signal issu de la caméra 123 et fournit des informations d'expressions de visage, telle que la fermeture des yeux, les sourires et les froncements de front ou de sourcils, en mettant en oeuvre des algorithmes de traitements d'images connus. A titre d'exemple, on
25 analyse les contrastes de différentes zones du visage pour déterminer si des plis sont apparus sur la peau de chacune de ces zones, ce qui permet de détecter les sourires, et les expressions basées sur les plis frontaux. De même, en recherchant une zone colorée relativement circulaire, dans
30 une zone particulière du visage, on peut déterminer si les yeux sont ouverts et quelle est leur position. Enfin,

l'orientation de la tête peut être aussi déterminée. On pourra, de nouveau, se référer aux actes et proceedings du salon IMAGINA 1997, aux pages 246 à 257, pour la description de différents traitements d'images applicables
5 aux images vidéos transmises par la caméra 123.

La console numérique 125 est actionnée par un technicien et fournit une information représentative de l'animateur et de ses invités, ces différents interlocuteurs étant associés chacun à un ou plusieurs
10 microphones. Ainsi le signal sonore issu de chaque microphone est automatiquement associé à un signal représentatif de l'identité de l'interlocuteur qui s'exprime. A chaque instant, seul l'un des interlocuteur est ainsi identifié par le signal sortant de la console
15 numérique 125 ou du moyen de traitement de sons 133.

L'information d'expression visuelle déterminée par le moyen de traitement d'image 123 et les signaux issus de la console numérique 125 et du moyen de traitement de sons 133, sont associés au signal sonore stéréophonique,
20 par le moyen d'insertion de données numériques 134, selon des techniques connues dans le domaine de la transmission d'informations alphanumériques sur un canal radiophonique, par exemple par modulation de la sous-porteuse du signal véhiculé sur ce canal.

25 On comprend aisément que les signaux émis par l'émetteur 126 à destination d'une multitude de récepteurs 127 sont simultanément représentatifs de :

- la voix de chaque interlocuteur qui s'exprime, par l'intermédiaire des microphones, en stéréophonie,

- l'identité de cet interlocuteur, par l'intermédiaire de la console numérique 125 ou du moyen de traitement de sons 133,
- la position de la tête de cet interlocuteur, par l'intermédiaire du moyen de traitement de sons 133, et
- des expressions visuelles de cet interlocuteur, par l'intermédiaire du moyen de traitement d'image 124.

Chaque récepteur 127 comporte une antenne hertzienne et est adapté à recevoir le signal diffusé par l'émetteur hertzien 134. Ce signal est démodulé par le récepteur 127 et transmis au moyen d'analyse de signaux sonores 128 qui analyse les signaux sonores et fournit des informations représentatives d'expressions du visage de l'interlocuteur, de phonèmes prononcés, de position de tête et d'expressions faciales.

Le moyen de mémorisation de modèle est ici un compact disque 150 associé à un lecteur de compact disque, par exemple de type mémoire morte connu sous le nom de CD-ROM. Ce CD-ROM conserve des données d'images représentatives de modèles correspondant à un grand nombre d'animateurs et d'invités possibles et des points caractéristiques qui permettent d'animer ces visages modélisés.

Le moyen d'animation de modèle 129, constitué ici d'un ordinateur qui réalise conjointement les fonctions du moyen d'analyse de signaux sonores 128, anime le visage modélisé de l'interlocuteur, qui est conservé dans le moyen de mémorisation de modèle 109 et qui correspond au signal associé par la console numérique 125 à chaque microphone. Le moyen d'animation de modèle donné à ce visage modélisé :

- les expressions de visage fournies par le moyen d'analyse de signaux sonores 128,
- la position de la tête de cet interlocuteur, et
- les expression visuelles de cet interlocuteur.

5 l'image animée et sonorisée de l'interlocuteur considéré est diffusée :

- par l'intermédiaire d'un haut-parleur 131, pour la voix de l'interlocuteur, éventuellement combinée à d'autres signaux sonores caractéristiques de l'émission,
- 10 et
- par l'intermédiaire de l'écran de visualisation 132, pour l'image visage animé de l'interlocuteur, éventuellement complété par un décor, des personnages et une mise en scène caractéristiques de l'émission.

15 En figure 3, le troisième mode de réalisation de la présente invention est mis en oeuvre dans un réseau informatique et anime un visage par des expressions faciales qui correspondent à un texte transmis à distance.

Chaque ordinateur 141 relié à ce réseau

20 comporte ici un modem 144 et un moyen de mémorisation de modèle 142. L'ordinateur 141 est associé à un écran de visualisation 141 et à un clavier 143. L'ordinateur 141 est de type connu, par exemple utilisant un processeur PENTIUM (marque déposée) de la société INTEL (marque déposée). Le

25 modem 144 est de type connu. Il est adapté à transmettre sur un réseau téléphonique quelconque des données numériques. Le moyen de mémorisation de modèle 142 est ici constitué d'un disque dur sur lequel sont enregistrées des informations graphiques représentant le modèle destiné à

30 être animé ainsi que les points caractéristiques destinés à son animation.

Par l'utilisation du clavier 143 et de l'écran de visualisation 141, l'interlocuteur entre en mémoire de l'ordinateur 140 un texte. A tout ou partie des mots de ce texte sont associés des expressions faciales particulières, des mouvements de visage ou de corps et un décor choisi
5 parmi une multitude de décors, en mettant en oeuvre un logiciel.

Ainsi, il peut associer un sourire à un mot particulier, un clignement d'oeil à un autre mot et un
10 hochement de tête à un troisième mot. A cet effet, ledit logiciel présente le texte considéré sur des lignes parallèles à des ligne dite « d'expression » ou l'interlocuteur peut positionner des icônes représentatifs d'expressions, de mouvements, de décors, de figures
15 graphiques ou d'images numériques, en regard de chaque mot. On observe que ces indications suffisent à l'homme du métier de l'informatique pour qu'au fichier texte considéré soit associé des données représentatives des icônes positionnés par l'interlocuteur. En conséquence, ce
20 logiciel n'est pas plus détaillé ici.

Le modem 144 module, sur la prise téléphonique 145, un signal sonore dans les fréquences de la bande passante d'une ligne téléphonique, pour que ce signal représente :

- 25 - le modèle du visage de l'interlocuteur,
- le texte susceptible d'être dit par cet interlocuteur et
- les informations accessoires destinées à l'animation de modèle.

30 Un modem 150, relié par l'intermédiaire d'une ligne téléphonique 146 et d'une prise téléphonique 147 au

modem émetteur 144, reçoit ce signal et restitue les fichiers correspondant aux trois types d'informations mentionnés au paragraphe précédent.

Le modem 150 est, dans le troisième mode de réalisation, incorporé dans chaque ordinateur 148 relié audit réseau. Chaque ordinateur 148 constitue, en outre, un moyen d'analyse de messages 153 et un moyen d'animation de modèle 152. Cet ordinateur 148 est associé, de manière connu, à un écran de visualisation 149, à un haut-parleur 154 et à un clavier 151.

Le moyen d'analyse de message 153 est constitué du processeur de l'ordinateur 148 et d'un logiciel d'analyse de données textuelles de type connu, qui associe à ce texte une suite de phonèmes correspondant à la prononciation de ce texte.

Le moyen d'animation de modèle 152, constitué ici de la mise en oeuvre du processeur de l'ordinateur 148, et d'un logiciel d'animation approprié, associe à chaque phonème fourni par le moyen d'analyse de message 153, une expression faciale correspondant à la prononciation de ce phonème, selon des techniques par exemple rappelées dans les documents de l'art antérieur cités dans le préambule de la présente demande, et dans les documents auxquels ils font références qui sont tous conjointement incorporés ici par référence.

Le moyen d'animation de modèle 152 anime le visage modélisé de l'interlocuteur, tel qu'il est reçu par l'intermédiaire du modem 150, pour donner à ce visage modélisé :

- les expressions de visage correspondant à la prononciation des phonèmes,

- les expressions accessoires associées au texte par l'interlocuteur, et
- les éléments de décors et d'environnement associés au texte par l'interlocuteur.

5 En outre, le moyen d'animation de modèle 152 fournit des données sonores à une carte d'émission de sons, selon des techniques connues dans les systèmes de synthèse de voix ou de lecture automatique de textes, comme, par exemple, les standards téléphoniques électroniques
10 interactifs à synthèse de voix.

 l'image animée et sonorisée de l'interlocuteur considéré est diffusée :

- par l'intermédiaire d'un haut-parleur 154, pour la voix de l'interlocuteur, éventuellement combinée à
15 d'autres signaux sonores caractéristiques de choix de l'interlocuteur, comme l'accent de prononciation, les paroles prononcées à voix basse, et
- par l'intermédiaire de l'écran de visualisation 132, pour l'image visage animé de l'interlocuteur,
20 éventuellement complété par un décor, des personnages et une mise en scène caractéristiques de choix de l'interlocuteur rappelés ci-dessus.

 En figure 4, le quatrième mode de réalisation est mis en oeuvre dans un réseau informatique.

25 On retrouve en figure 4 les éléments de la figure 3, à l'exception du moyen d'analyse de messages 153 qui est remplacé par un moyen d'analyse de signaux sonores 160. L'ordinateur 140 comporte, en outre, un moyen de capture de sons 161, connu sous le nom de carte de
30 numérisation de sons, associé à un microphone 162, de type connu.

Le quatrième mode de réalisation fonctionne de la même manière que le troisième mode de réalisation illustré en figure 3, à la différence près que le fichier de données textuelles est remplacé par un fichiers de données sonores représentant la voix de l'interlocuteur. Comme dans le troisième mode de réalisation, l'interlocuteur associe à ce fichier des données d'animation, d'expression, de mouvement, de décors ou d'image.

10 Pour le destinataire de ce fichier, le moyen d'analyse de signaux sonores 160, constitué du processeur de l'ordinateur 148 et d'un logiciel dédié, analyse les signaux reçus par l'intermédiaire du modem 150, et fournit des informations représentatives d'expressions du visage de
15 l'interlocuteur correspondant à la prononciation du message vocal porté par lesdits signaux sonores.

Dans le schéma bloc de dispositif d'émission d'images mis en oeuvre dans le premier mode de réalisation de la présente invention (figure 5), sont représentés :

- 20 - un appareil téléphonique 1, comportant un combiné 2 possédant un microphone 30, et un clavier 3, et adapté à émettre un signal électrique représentatif des sons qui atteignent le microphone 30 et des pressions effectuées sur les touches du clavier 3, selon des techniques
25 connues dans les appareils téléphoniques à fréquences vocales ;
- un réseau téléphonique 4, de type connu, représenté sous la forme de deux rectangles schématisant des prises téléphoniques séparés par une ligne interrompue ;
- 30 - un moyen d'analyse de voix 5 dont le fonctionnement est exposé en regard de la figure 6 et qui fournit des

données représentatives d'expressions orales à un moyen d'animation de visage 9 ;

5 - un moyen d'analyse de fréquences vocales synthétiques 6, de type connu, adapté à transmettre des données numériques représentatives des fréquences vocales émises sur le réseau téléphonique par l'appareil téléphonique 1, et représentative d'une éventuelle pression appliquée sur un des touches de son clavier 3 ;

10 - un moyen de numérisation d'image 7, adapté à transmettre des données représentatives d'une image d'un visage, de type connu dans les scanners bureautiques ;

15 - un moyen de modélisation 8 adapté à combiner des données représentatives d'une image reçue de la part du moyen de numérisation d'image 7 avec un modèle et à fournir l'image fixe combinée résultante au moyen d'animation d'image 9 ;

20 - le moyen d'animation de visage 9 adapté à animer le modèle combiné avec des expressions faciales correspondant aux expressions orales provenant du moyen d'analyse de voix 5 ;

25 - un moyen d'animation de personne 10, de type connu, et adapté à fournir des informations représentatives d'images animées d'une personne possédant le visage réalisé par le moyen d'animation de visage 9, en fonction de données provenant du moyen d'analyse de fréquences vocales synthétiques 6 ;

30 - un moyen de combinaison de personne et de scène 11 adapté à insérer l'image de la personne représentée par les informations sortant du moyen d'animation de personne 10 avec des informations sur une scène

comportant, par exemple des mobiles, des décors et des personnages, selon des techniques connues ;

- un moyen d'affichage 12 de type connu, par exemple constitué d'un téléviseur, adapté à afficher l'image sortant du moyen de combinaison 11 ;

- un émetteur vidéo 13 de type connu, par exemple constitué d'une émetteur hertzien ou d'un émetteur sur réseau de télévision câblé ; et

- un moyen d'enregistrement 14 de type connu, par exemple constitué d'un magnétoscope.

Le moyen d'analyse de voix 5 qui fournit des données représentatives d'expressions orales est, par exemple, constitué d'un ordinateur, d'une carte d'acquisition de son, de type connu, et d'un logiciel de détection dit "F.V.", initiales des mots "Fréquences Vocales", qui détermine des clés d'animation, en fonction des fréquences vocales utilisées. Ainsi, en fonction des graves, des aigus, des paliers et des évolutions de ces fréquences utilisées, sont déterminés des phonèmes et des lettres prononcés, des rires, des attentes sonores ou muettes, ... et d'autres expressions et attitudes faciales. Des fournisseurs connus de ces types de logiciels sont SILICLONE et SOFTIMAGE, VIEWER. On observe que, préférentiellement, une étape initiale d'apprentissage est effectuée avec l'interlocuteur et hors antenne. Au cours de cette étape, on demande à l'interlocuteur de prononcer un certain nombre de phonèmes, par exemple les voyelles et on mémorise les spectres caractéristiques des prononciations de ces phonèmes.

Le moyen de modélisation 8 est adapté à combiner des données représentatives d'une image reçue de

la part du moyen de numérisation d'image 7 avec un modèle. Il est, par exemple, composé d'un ordinateur et d'un logiciel de traitement d'image adapté à réaliser des traitements d'images connus sous le nom anglais de
5 "morphing", parfois traduit en français sous le nom de "métamorphose" et qui fait correspondre :

- des points caractéristiques de deux images (coins de bouche, points de contours du nez et du visage, coins des yeux) pour déterminer la distance entre ces points
10 sur les deux images (ici l'une des images est un modèle de visage en trois dimensions et l'autre est une photographie numérisée) ; et
- les autres points, par propagation et atténuation progressive de ces distances, en calculant sur des
15 triangles qui entourent ces points et dont les sommets sont des points caractéristiques, un déplacement du point de la photographie.

Certains ouvrages techniques de référence, comme le guide de l'utilisateur et le manuel de
20 l'utilisateur du logiciel SOFTIMAGE "EDDIE", imprimé au Canada, définissent le morphing ou la métamorphose comme un fondu enchaîné entre deux états. Ils sont incorporés à la description par référence.

On note que, selon l'invention, le visage
25 modélisé résultant du fonctionnement du moyen de modélisation 8 correspond à un état intermédiaire de la métamorphose entre une photographie d'un visage réel et d'un modèle destiné à être animé.

Ainsi, le visage modélisé résultant possède un
30 visage de l'interlocuteur représenté sur la photographie, les traits de ce visage pouvant être animés par déplacement

de points ou de lignes particuliers du modèle auquel a été associée la photographie pour former le visage modélisé.

Le moyen d'animation de visage 9 adapté à animer le modèle combiné avec des expressions faciales
5 correspondant aux expressions orales provenant du moyen d'analyse de voix 5 est de type connu dans l'animation de personnages de dessins animés. Le lecteur pourra, pour une meilleure compréhension, se référer aux ouvrages de référence cités ci-dessus ainsi qu'au "Tool Book" et au
10 manuel de l'utilisateur du logiciel correspondant de la société ALIAS WAVE FRONT et du guide de l'utilisateur du logiciel "Morph" de la société GRYPHON SOFTWARE CORPORATION, ces cinq documents étant incorporés par référence dans la description de l'invention.

15 L'ensemble des fonctions de traitement d'image décrites en regard des figures 5 et 6 peuvent être réalisées par des ordinateurs mettant en oeuvre des logiciels tels que ceux cités ci-dessus ainsi que les logiciels de marque SOFTIMAGE VIEWER, SILICLONE et un
20 logiciel de l'Institut National de l'Audiovisuel réalisant des morphings ou métamorphoses en trois dimensions de manière automatique.

Le fonctionnement du dispositif présenté en figure 5 est simple : pour transmettre une image animée, le
25 dispositif utilise d'une part un traitement d'une image fixe afin de la rendre apte à être animée (par l'intermédiaire du moyen de numérisation d'image 7 et du moyen de modélisation 8) et d'autre part un traitement de voix pour déterminer des expressions orales (réalisée par
30 le moyen d'analyse de voix 5) puis faciales qui animent

l'image fixe (animation effectuée par le moyen d'animation de visage 9).

Selon une variante, destinée à la transmission d'images entre interlocuteurs téléphoniques, l'appareil
5 téléphonique peut ne pas être à fréquence vocale, le moyen d'analyse de fréquences vocales 6, le moyen d'animation de personne 10, le moyen de combinaison de personne et de scène 11, l'émetteur vidéo 13 et le moyen d'enregistrement 14 sont supprimés, le moyen d'affichage 12 affichant
10 directement l'image sortant du moyen d'animation de visage.

La figure 6 représente les opérations successives réalisées par le dispositif illustré en figure 5, en mettant en oeuvre un programme conservé dans la mémoire morte 17 de l'ordinateur 15 (figure 7).

15 Les deux premières opérations, référencées 200 et 201, sont effectuées préliminairement à la réception du signal sonore qui porte la voix. Au cours de l'opération 200, le visage est numérisé, par le moyen de numérisation 7 (figure 5) et mémorisé dans la mémoire vive 16 de
20 l'ordinateur 15 (figure 7).

Au cours de l'opération 201, le visage numérisé au cours de l'opération 200 est combiné, par métamorphose, avec un modèle destiné à être animé. Les traits du visage résultant sont donc ceux du visage numérisé, c'est-à-dire
25 ceux de l'interlocuteur, mais les éléments de ce visage sont mis en mouvement selon des procédures dépendant du visage numérisé, mais liées aussi au modèle destiné à être animé.

Le visage résultant est donc hybride, son
30 apparence étant celle du visage numérisé et ses mouvements étant ceux du modèle. Comme il est décrit en regard de la

figure 5, les mouvements sont commandés, par l'intermédiaire de l'analyse de la voix de l'interlocuteur, par les mouvements du visage de l'interlocuteur.

L'opération 202 correspond à la réception du son téléphonique sur une entrée de signaux du dispositif présenté en figure 5.

L'opération 203 correspond à l'analyse spectrale du son reçu au cours de l'opération 202 afin de fournir un spectre en fréquence de ce signal. Dans ledit spectre, à chaque fréquence, en abscisse, est associée l'amplitude du signal possédant cette fréquence, en ordonnée, dans le signal sonore reçu (figure 10).

L'opération 204 consiste à extraire les fréquences vocales synthétiques pour déterminer si le clavier 3 de l'appareil téléphonique 1 a été utilisé. On note que les fréquences vocales utilisées correspondent toujours à une combinaison d'au moins deux fréquences primaires non harmoniques, de telle manière que les risques d'erreur de détection soient limités.

Le résultat de l'extraction des fréquences vocales, est la mise en mémoire vive 16, dans le registre *fregvoc* de la valeur de la touche du clavier sur laquelle une pression a été détectée (figure 7).

Au cours de l'opération 205, le moyen d'analyse de voix détermine, par analyse du spectre réalisé au cours de l'opération 203, les phonèmes et autres expressions orales utilisées par l'interlocuteur. Pour réaliser cette opération, le spectre du signal sonore est comparé à des spectres caractéristiques de phonèmes et d'expressions orales (comme le rire), lesdits spectres caractéristiques étant conservés dans la mémoire morte 17 de l'ordinateur

15. On note cependant, que cette analyse est effectuée de manière dynamique, ce qui signifie que ce n'est pas toujours un seul spectre instantané qui permet de déterminer l'expression orale mais aussi parfois une
5 succession de spectres, ladite succession pouvant être caractéristique d'une expression orale. Pour une meilleure compréhension de cette opération, on se rapportera aux notices d'utilisation des logiciels cités plus haut.

Au cours de l'opération 206, chaque expression
10 orale est mise en relation avec une expression faciale, par exemple le phonème prononcé à la lecture de la lettre "O" correspond à une expression faciale dans laquelle les mâchoires sont légèrement écartées et les lèvres forment une ouverture ronde de petite dimension.

15 Chacune de ces expressions faciales est mise en correspondance, au cours de l'opération 207, avec une succession de mouvements de points caractéristiques du modèle de visage qui supporte le visage de l'interlocuteur. Par exemple si la prononciation de la lettre "O" est
20 effectuée pendant une durée d'une seconde, entre deux silences, les mouvements successifs du modèle et donc de la personne modélisée, correspondent :

- d'abord, pendant un quart de seconde, à l'écartement des mâchoires et à la mise des lèvres en rond ;
- 25 - puis, à un maintien de cette posture pendant toute la durée de la prononciation de la lettre "O" ;
- puis, à la fin de cette prononciation (et donc ici au retour du silence), le visage reprend son aspect initial en un quart de seconde.

30 Si la fréquence de rafraîchissement de l'image est de 25 Hertz, c'est-à-dire que 25 images différentes

sont affichées pendant chaque seconde (fréquence correspondant à celle de la télévision), la séquence décrite ci-dessus correspond à environ 37 images donc à 37 successions de mouvements effectués entre deux images.

5 L'opération 208 consiste à mettre le visage en mouvement selon les valeurs successives conservées dans le registre *fregvoc* (figure 7) pour tenir compte de l'appui sur les touches du clavier 3. Cette opération 208 correspond par exemple à un jeu auquel joue
10 l'interlocuteur.

L'opération 209 consiste à insérer le visage résultant des opérations 207 et 208 dans une scène dont les caractéristiques dépendent, elles aussi, des expressions orales et des touches de clavier utilisées.

15 Par exemple, au cours d'un jeu, les expressions orales peuvent permettre de déformer un objet, et les appuis sur des touches peuvent permettre de déplacer cet objet.

L'opération 210 correspond à l'affichage de la
20 scène comportant le visage, à sa mémorisation et à l'émission de l'image. Dans l'exemple du jeu télévisé considéré ici, l'interlocuteur voit son image sur un téléviseur et il utilise, d'une part le clavier 3 de son appareil téléphonique 1, et d'autre part sa voix, pour
25 jouer au jeu considéré en voyant l'image animée de son visage sur son écran de télévision.

A la suite de l'opération 210, le système retourne à l'opération 202.

Il faut noter que la suite d'opérations décrite
30 ici s'effectue, en fait, en temps réel, c'est-à-dire que, après l'affichage de chaque image (opération 210), toutes

les opérations 202 à 209 sont effectuées en une durée inférieure à celle du rafraîchissement de l'image, le visage et la scène étant en fait modifiés progressivement au cours de ces opérations.

5 La figure 7 représente une implantation matérielle d'un dispositif présenté en figure 5. Ce dispositif est, ici, organisé selon des architectures connues dans le domaine des ordinateurs, autour d'un bus de communication informatique 20 auxquels sont reliés :

- 10 - une unité centrale de calcul 21, de type connu, comportant en particulier un processeur ;
- une mémoire vive 16 qui comporte des registres de mémoire et en particulier le registre *fregvoc* ;
- une mémoire morte 17 qui conserve le programme de
- 15 fonctionnement du dispositif ;
- un port de sortie vidéo 18 ;
- un port d'entrée et de traitement de son 19 ; et
- un port d'entrée et de traitement d'image 22.

 Le port de sortie vidéo 18 est de type connu,

20 il fournit, selon des variantes, soit un signal adapté aux normes en matière de télévision, par exemple la norme CCIR en Europe, soit un signal adapté aux normes en matière d'écrans d'ordinateur, par exemple la norme super VGA.

 Le port d'entrée et de traitement de son 19 est

25 de type connu. D'une part il numérise le son, et d'autre part il en extrait un spectre instantané (figure 10). Les informations résultant de ce traitement sont mémorisées dans la mémoire vive 16.

 Le port d'entrée et de traitement d'image 22

30 est de type connu. Il permet de mettre en mémoire vive 16

des données numériques représentatives d'une scène visuelle.

La figure 8 représente une photographie d'un visage d'enfant. On observe que cet enfant a des cheveux
5 plats, des lunettes dont chaque verre est sensiblement plus large dans la partie haute que dans la partie basse, qu'il a les oreilles normalement collées et qu'il arbore un sourire tout en conservant la bouche fermée.

La figure 9 représente un modèle du visage
10 présenté en figure 8, auquel un microphone à bras articulé a été ajouté. On observe ici aussi que ce modèle d'enfant présente des cheveux plats, des lunettes dont chaque verre est sensiblement plus large dans la partie haute que dans la partie basse, qu'il a les oreilles normalement collées
15 et qu'il arbore un sourire tout en conservant la bouche fermée.

Le visage modélisé créé par le moyen de modélisation 8 au cours de l'opération 201 fournit une image qui est intermédiaire dans la métamorphose des images
20 présentées en figure 8 et 9. On observe (figure 11) que les visages des figures 8 et 9 peuvent être associés par des points caractéristiques 60, sur le front, 61, sur les pommettes, 62, aux coins des lèvres et 63 sur le menton. Ces points nommés "primaires" sont ceux qui sont les
25 sources ou repères de déformation du visage, c'est-à-dire que les autres points du visage sont déplacés en fonction du déplacement de ces points primaires, pour représenter l'élasticité de la peau du visage.

Bien que dès à présent, le positionnement
30 automatique des points primaires sur l'image représentée en figure 8 soit possible, par des techniques de

reconnaissance de forme appliquées à ces points (voir en particulier les proceedings du salon IMAGINA sus-mentionnés), il est plus économique et plus sûr de positionner manuellement, en utilisant des moyens de pointage connus, tels que souris ou joystick, associés à l'unité centrale 21 pour mettre en oeuvre le procédé de l'invention dans le cadre d'émissions télévisuelles.

C'est le déplacement des points primaires qui est mis en relation avec des expressions faciales, elles-mêmes en relation avec des expressions orales analysées à partir de la voix de l'interlocuteur.

La figure 10 représente le spectre d'un signal sonore représentatif d'une voix.

Comme indiqué plus haut, en abscisse de ce spectre se trouvent les fréquences et en ordonnée les amplitudes correspondantes. C'est la répartition spectrale et la déformation, dans le temps, de cette répartition qui permettent de déterminer une expression orale.

La figure 11 représente l'expression qui correspond au signal sonore présenté en figure 10.

Cette expression de rire comporte, par rapport au même visage sans expression :

- un déplacement vers la bas du point primaire du menton 63 ;
- un écartement et une remontée latérale des points primaires des lèvres 62 ;
- une remontée des points primaires de pommettes 61 ;
- un rapprochement des points primaires de front 60.

La figure 12 représente l'image animée correspondant à l'expression portée par la voix présentée en figure 10. Chaque point du visage modélisé, représenté

en figure 9, est associé à trois points primaires, sommets d'un triangle qui entoure le point considéré. Ce point est déplacé proportionnellement à la déformation et au déplacement des sommets de ce triangle, de telle manière
5 que les déplacements soient continus sur toute la surface du triangle.

On observe que deux procédures peuvent indifféremment être suivies : soit, comme exposé ci-dessus, on constitue par métamorphose partielle, une personne
10 modélisée, dont certains points, dits primaires, servent de repères pour l'animation, soit on anime le modèle puis on constitue une métamorphose partielle du modèle animé et de l'image réelle.

Selon une variante non représentée, le
15 modéliseur 8 ne fonctionne plus à partir d'une photographie, mais à partir d'informations transmises par l'interlocuteur utilisant le téléphone 2. A titre d'exemple, cet utilisateur indique les caractéristiques de son visage, de sa chevelure, de lunettes, d'appareil
20 dentaire, de couleur de peau, de maigreur et autres caractéristiques optiques. A cet effet, il peut utiliser de nombreuses techniques connues, par exemple en énonçant ces caractéristiques, en utilisant un ordinateur et un modem, en utilisant le clavier 3, en remplacement de l'information
25 fournie par le scanner 7.

Selon une dernière variante, le modéliseur 8 reçoit une information représentative d'une image d'une personne modélisée conservée en mémoire et une voix destinée à être associée à cette personne modélisée.
30 L'image mémorisée comporte les repères ou points primaires présentés ci-dessus. A titre d'exemple, l'information

représentative peut indiquer la personne politique ou médiatique qui est représentée et la voix peut être la voix de cette personne ou encore une imitation de la voix de cette personne.

5 On observe que la mise en mémoire des personnes modélisées adaptées à être animées par les voix peut être réalisée soit par transmission à distance, par exemple par le réseau téléphonique ou par une émission hertzienne, soit par distribution de supports de mémoire, comme des disques
10 optiques ou magnéto-optique ou des compact-disques (CD-ROM).

 On note à cet égard que l'invention permet ainsi la création de chaînes de télévision d'information ou de divertissement en utilisant uniquement la bande passante
15 correspondant aux voix humaines, éventuellement complétée par des informations numériques traitées à leur réception pour ne pas être audibles par les spectateurs. Ainsi, une station de radiodiffusion pourra ajouter à ses programmes normaux des informations qui, traitées à la réception,
20 permettront au dispositif selon l'invention de déterminer la personne modélisée à animer par la voix transmise par ladite station (cette personne modélisée pouvant, en complément être animée par des informations numériques non audibles transmises par le signal radiodiffusé).

25 De mêmes des utilisateurs de réseaux de communication, comme le réseau INTERNET, pourront non seulement transmettre leur image et l'animer avec leur voix, mais aussi transmettre des images et des voix d'autres personnes.

Selon une variante, l'image de l'interlocuteur est prise par une caméra vidéo réaliée à une carte de numérisation d'image.

Selon une variante non représentée de chaque
5 mode de réalisation, le moyen d'analyse de message ou de sons n'effectue que la mesure d'une intensité sonore instantanée et le moyen d'animation de modèle n'effectue que l'animation de la bouche du visage modélisé, en ouvrant la bouche par séparation verticale des lèvres, avec une
10 ouverture des lèvres d'autant plus grande que l'intensité sonore instantanée analysée par le moyen d'analyse est élevée.

Selon une dernière variante, l'interlocuteur dont l'image est transmise, transmet son image
15 électronique, captée par un capteur photoélectrique connu, en positionnant lui-même les points primaires ou repères permettant d'animer cette image en utilisant les expressions orales transmises par la voix, comme décrit ci-dessus. Pour positionner lesdits repères, un système
20 informatique comportant un ordinateur et un moyen de pointage, comme par exemple une souris, pourront être utilisés selon des techniques connues de l'homme du métier.

L'invention s'applique particulièrement aux transmissions de messages audio-visuels sur un réseau
25 informatique, de type INTERNET, à la diffusion d'émission télévisuelles ou radiophoniques, à l'incorporation dans des consoles de jeux, des ordinateurs, des bancs de montage audiovisuels ou encore des télévisiophones (non représentés).

REVENDEICATIONS

1. Dispositif de transmission d'images animées et sonorisées représentatives d'un interlocuteur, caractérisé en ce qu'il comporte :
- un moyen de mémorisation (109, 142) d'une image d'un visage dit « modélisé » représentatif du visage dudit interlocuteur ;
 - un moyen de communication à distance (101 à 105, 140 à 144) de signaux représentatifs d'un message susceptible d'être prononcé par ledit interlocuteur ;
 - un moyen d'analyse desdits signaux (107, 153) adapté à déterminer une succession d'expressions de visage correspondant à la prononciation dudit message par une voix, selon des règles prédéterminées prenant en compte lesdits signaux ;
 - un moyen d'animation (110, 152) adapté à animer l'image du visage modélisé de l'interlocuteur pour que ledit visage présente successivement chaque expression de ladite succession d'expressions de visage correspondant à la prononciation dudit message par ladite voix, et
 - un moyen d'affichage d'image et d'émission de sons vocaux (113, 114, 149, 154), adapté à émettre ladite voix et à afficher simultanément ledit visage modélisé présentant l'expression correspondant à chaque instant à la prononciation dudit message par ladite voix.
2. Dispositif de transmission d'images animées et sonorisées représentatives du visage d'un interlocuteur, caractérisé en ce qu'il comporte :

- un moyen de mémorisation (109) d'une image d'un visage dit « modélisé » représentatif du visage dudit interlocuteur ;
- un moyen de réception de signaux (106) représentatifs d'un message susceptible d'être prononcé par ledit interlocuteur ;
- un moyen d'analyse desdits signaux (107) adapté à déterminer une succession d'expressions de visage correspondant à la prononciation dudit message par une voix, selon des règles prédéterminées prenant en compte lesdits signaux ;
- un moyen d'animation (110) adapté à animer l'image du visage modélisé de l'interlocuteur pour que ledit visage présente successivement chaque expression de ladite succession d'expressions de visage correspondant à la prononciation dudit message par ladite voix, et
- un moyen de transmission (111) de signaux simultanément représentatifs de sons vocaux, correspondant à ladite voix et d'images correspondant audit visage modélisé présentant l'expression correspondant à chaque instant à la prononciation dudit message par ladite voix.

3. Dispositif de transmission selon l'une quelconque des revendications 1 ou 2, caractérisé en ce que le moyen de transmission à distance (101 à 105) est adapté à transmettre des signaux représentatifs d'un message vocal prononcé par ledit interlocuteur.

4. Dispositif de transmission selon l'une quelconque des revendications 1 ou 2, caractérisé en ce que

le moyen de transmission à distance (144) est adapté à transmettre des signaux représentatifs d'un texte susceptible d'être prononcé par l'interlocuteur.

5 5. Dispositif de transmission selon l'une quelconque des revendications 1 à 4, caractérisé en ce qu'il comporte un moyen de modélisation d'image (108, 7) adapté à fournir une image destinée à être animée, en fonction d'une image prise par un capteur photosensible et
10 en ce que le moyen de mémorisation (109, 8) mémorise ladite image.

6. Dispositif de transmission selon la revendication 5, caractérisé en ce que ledit moyen de
15 modélisation (109, 8) est adapté à effectuer un "morphing" partiel, c'est-à-dire une métamorphose partielle prenant en compte ladite image prise par un capteur photosensible, et un modèle prédéterminé adapté à être animé.

20 7. Dispositif de transmission selon l'une quelconque des revendications 1 à 6, caractérisé en ce qu'il comporte un récepteur (106, 147) relié à une ligne téléphonique (105, 146, 4).

25 8. Dispositif de transmission selon l'une quelconque des revendications 1 à 6, caractérisé en ce qu'il comporte une antenne hertzienne (111, 112, 126, 127, 13).

30 9. Réseau, caractérisé en ce qu'il comporte :

- une pluralité de systèmes informatiques reliés par des liaisons de communication,
- au moins un dispositif de transmission selon l'une quelconque des revendications 1 à 8, et
- 5 - relié par une dite liaison informatique à au moins un dit dispositif de transmission, au moins un moyen d'affichage et d'émission de sons vocaux, adapté à émettre ladite voix et à afficher simultanément ledit visage modélisé présentant l'expression correspondant à
- 10 chaque instant à la prononciation dudit message par ladite voix.

10. Procédé de formation d'images animées et sonorisées représentatives du visage d'un interlocuteur, 15 caractérisé en ce qu'il consiste à effectuer successivement les étapes suivantes :

- dans un premier temps :
 - * une opération de réception d'une image d'un visage dudit interlocuteur ;
 - 20 * une opération de modélisation d'un visage dit « modélisé » représentatif dudit visage et adapté à être animé ; et
 - * une opération de mémorisation dudit visage modélisé ;
- puis, dans un deuxième temps et après une opération de 25 réception de signaux représentatifs d'un message susceptible d'être prononcé par ledit interlocuteur,
 - * une opération d'analyse de ces signaux pour déterminer une succession d'expressions de visage qui correspond à la prononciation dudit message par une voix ; et,
 - 30 simultanément :

. une opération d'émission sonore de ladite voix prononçant ledit message, et

. une opération d'animation dudit visage modélisé pour qu'il présente successivement chaque expression de ladite succession d'expressions de visage
5 correspondant à chaque instant à la prononciation dudit message par ladite voix.

11. Procédé selon la revendication 10, caractérisé
10 en ce que l'opération de réception de signaux comporte la réception de signaux représentatifs d'un message vocal prononcé par ledit interlocuteur.

12. Procédé selon la revendication 11, caractérisé
15 en ce que l'opération de réception de signaux comporte la réception de signaux représentatifs d'un texte susceptible d'être prononcé par l'interlocuteur.

13. Procédé selon l'une quelconque des
20 revendications 10 à 12, caractérisé en ce que, au cours de l'opération de modélisation, ledit interlocuteur positionne des repères sur certains points d'une image captée par l'intermédiaire d'un capteur photosensible.

25 14. Console de jeu, caractérisée en ce qu'elle met en oeuvre un dispositif de transmission selon l'une quelconque des revendications 1 à 8 ou un procédé selon l'une quelconque des revendications 10 à 13.

30 15. Ordinateur, caractérisée en ce qu'il met en oeuvre un dispositif de transmission selon l'une quelconque

des revendications 1 à 8 ou un procédé selon l'une
quelconque des revendications 10 à 13.

16. Banc de montage audiovisuel, caractérisée en ce
5 qu'il met en oeuvre un dispositif de transmission selon
l'une quelconque des revendications 1 à 8 ou un procédé
selon l'une quelconque des revendications 10 à 13.

17. Télévisiophone, caractérisé en ce qu'il met en
10 oeuvre un dispositif de transmission selon l'une quelconque
des revendications 1 à 8 ou un procédé selon l'une
quelconque des revendications 10 à 13.

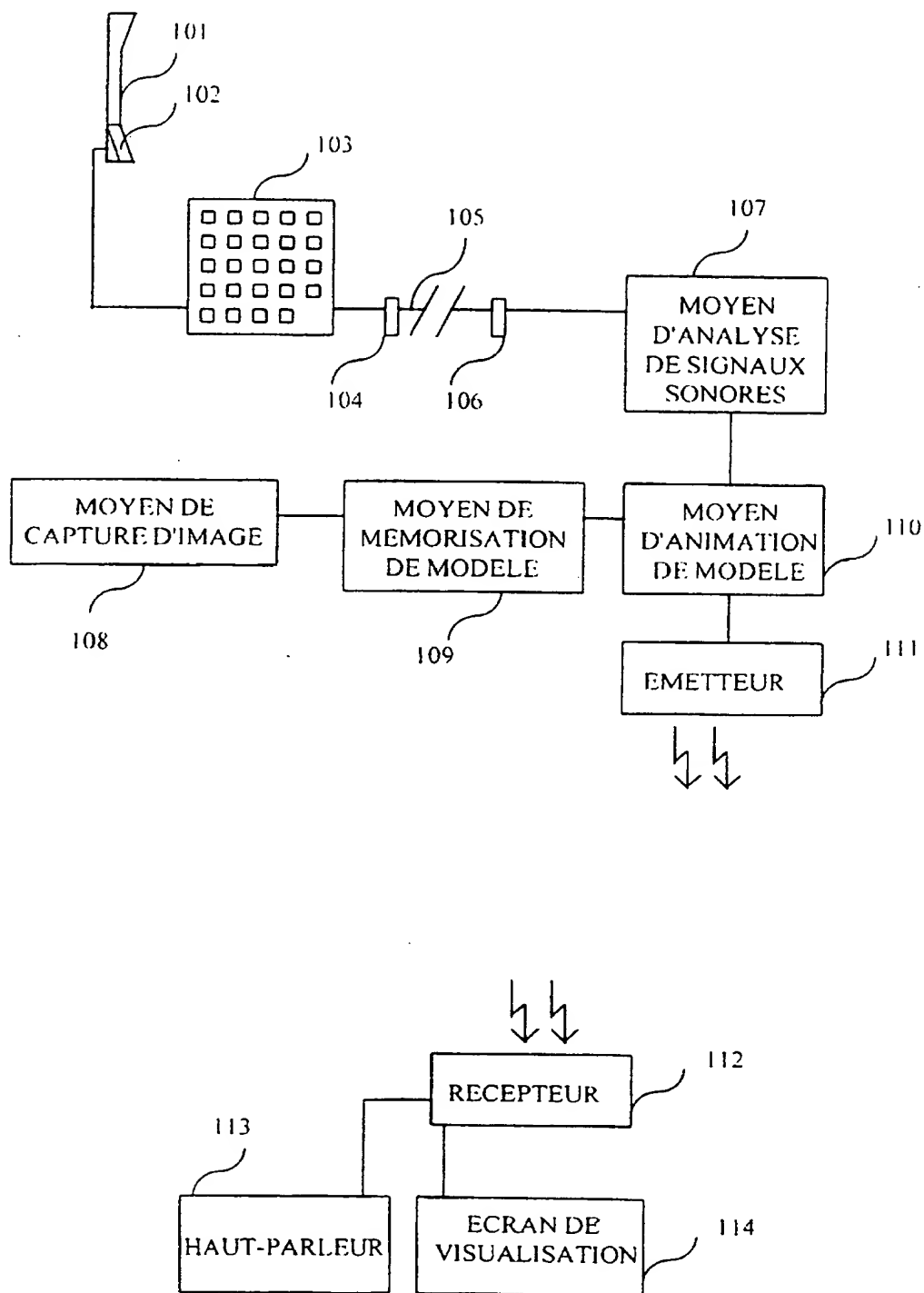


Fig. 1

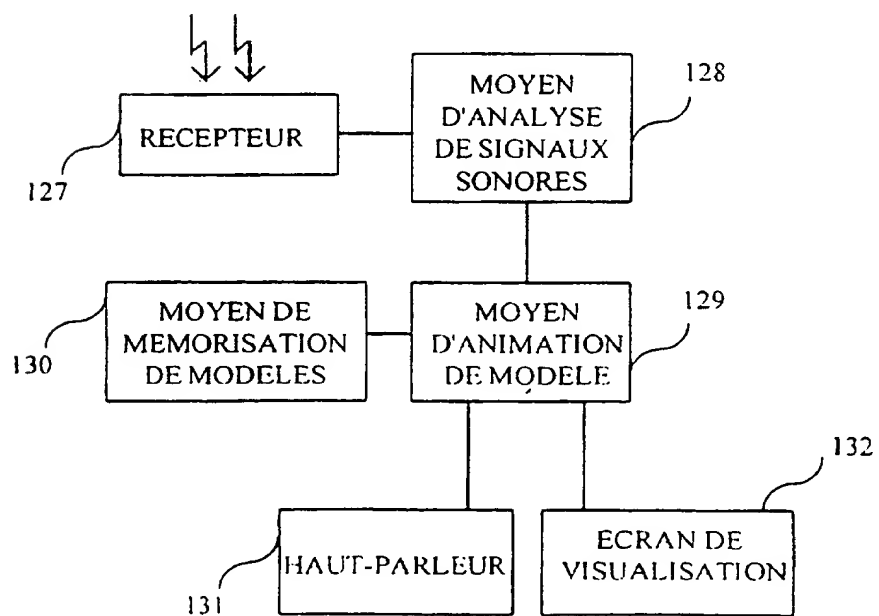
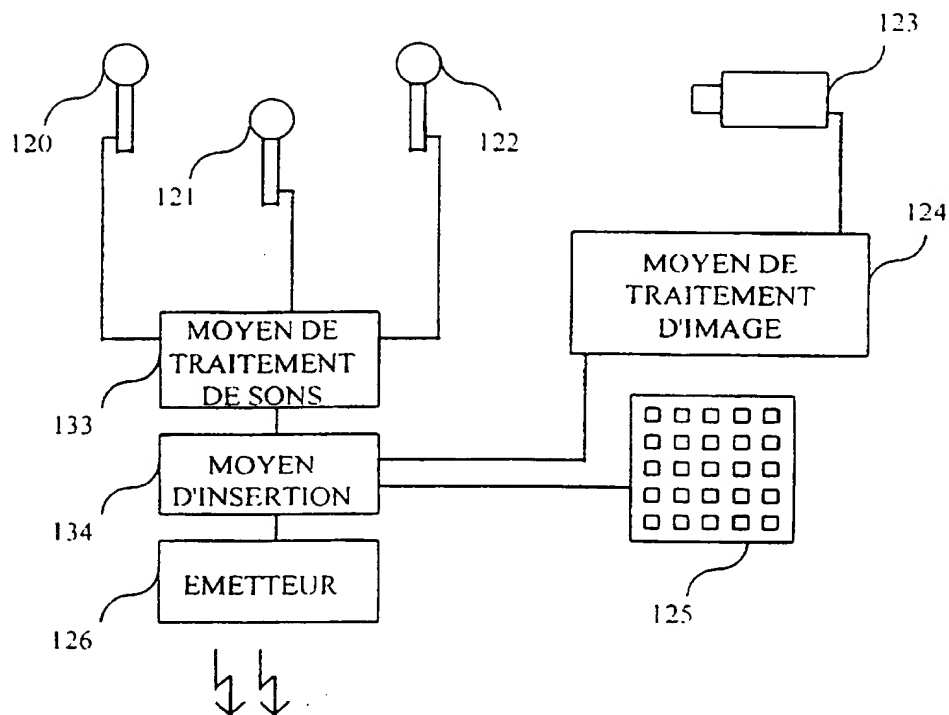


Fig. 2

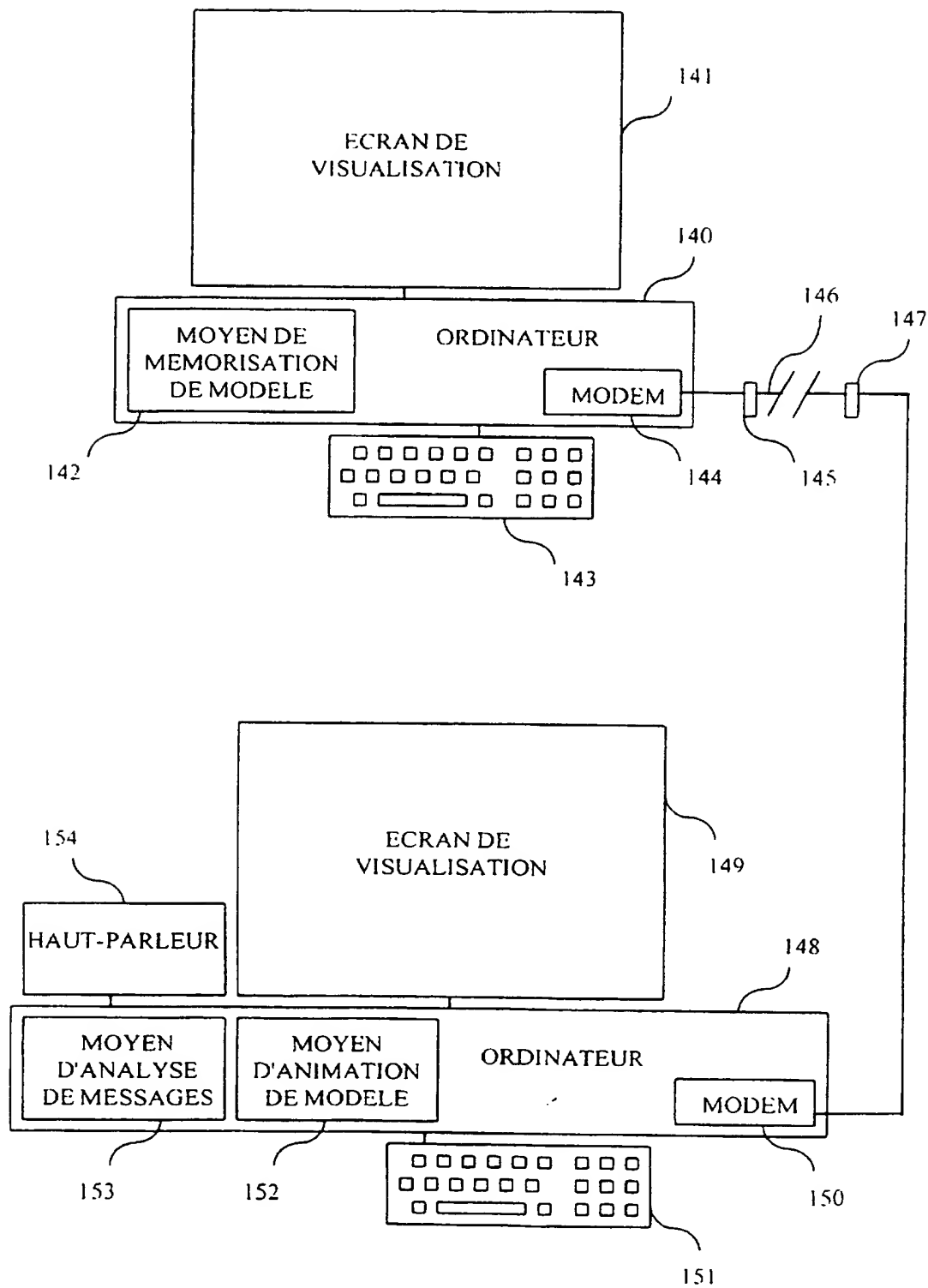


Fig. 3

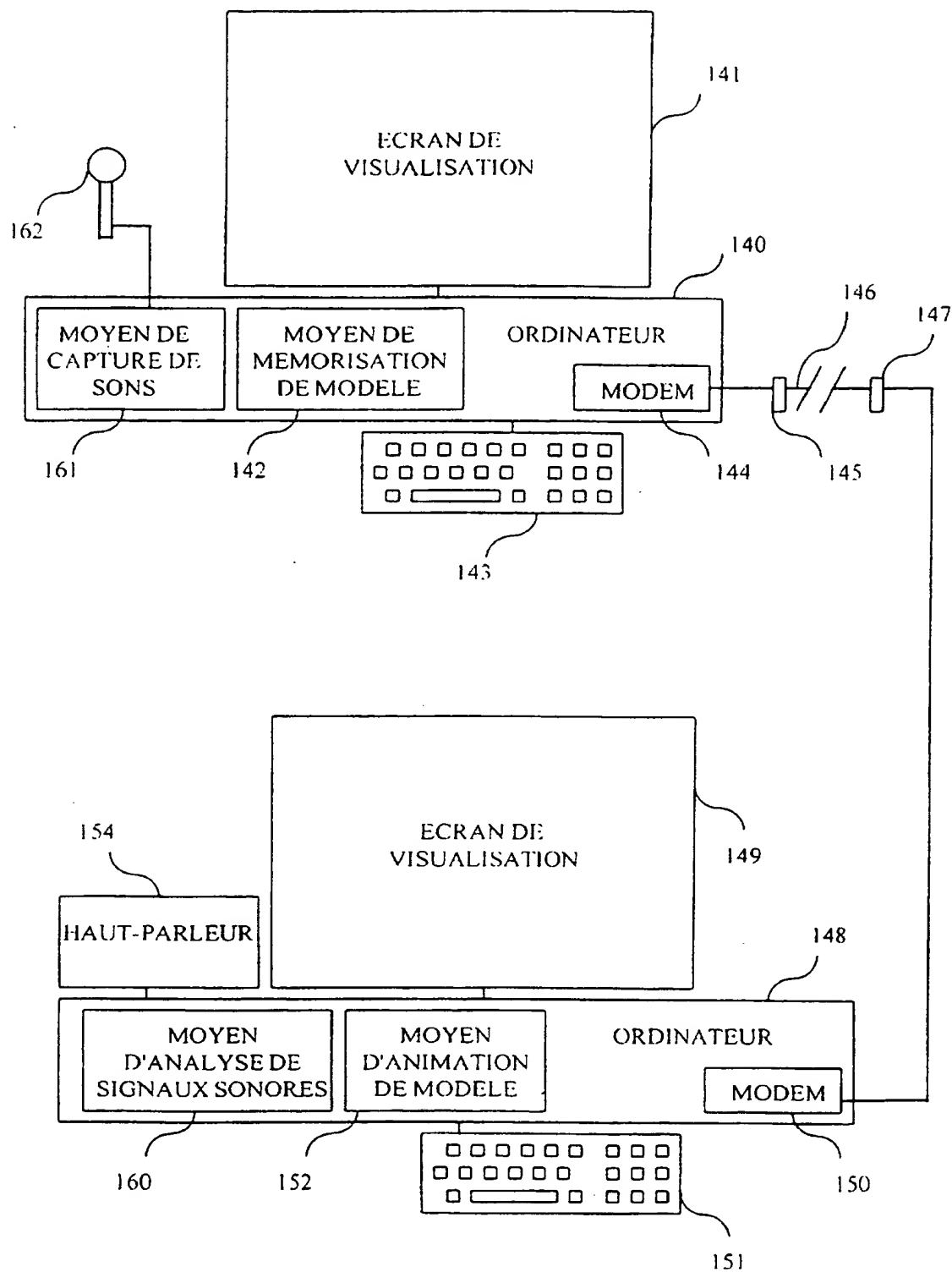


Fig. 4

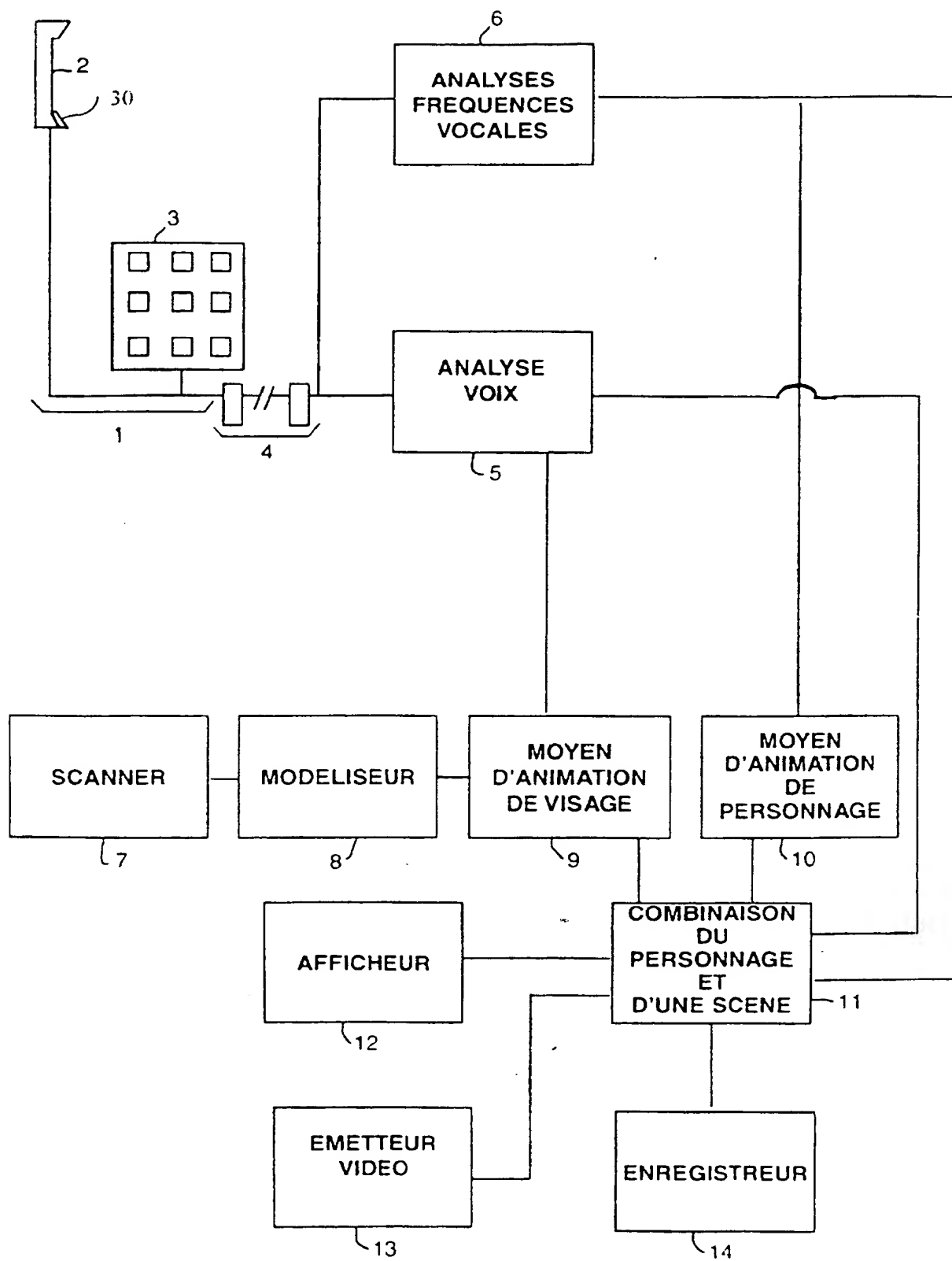


Fig. 5

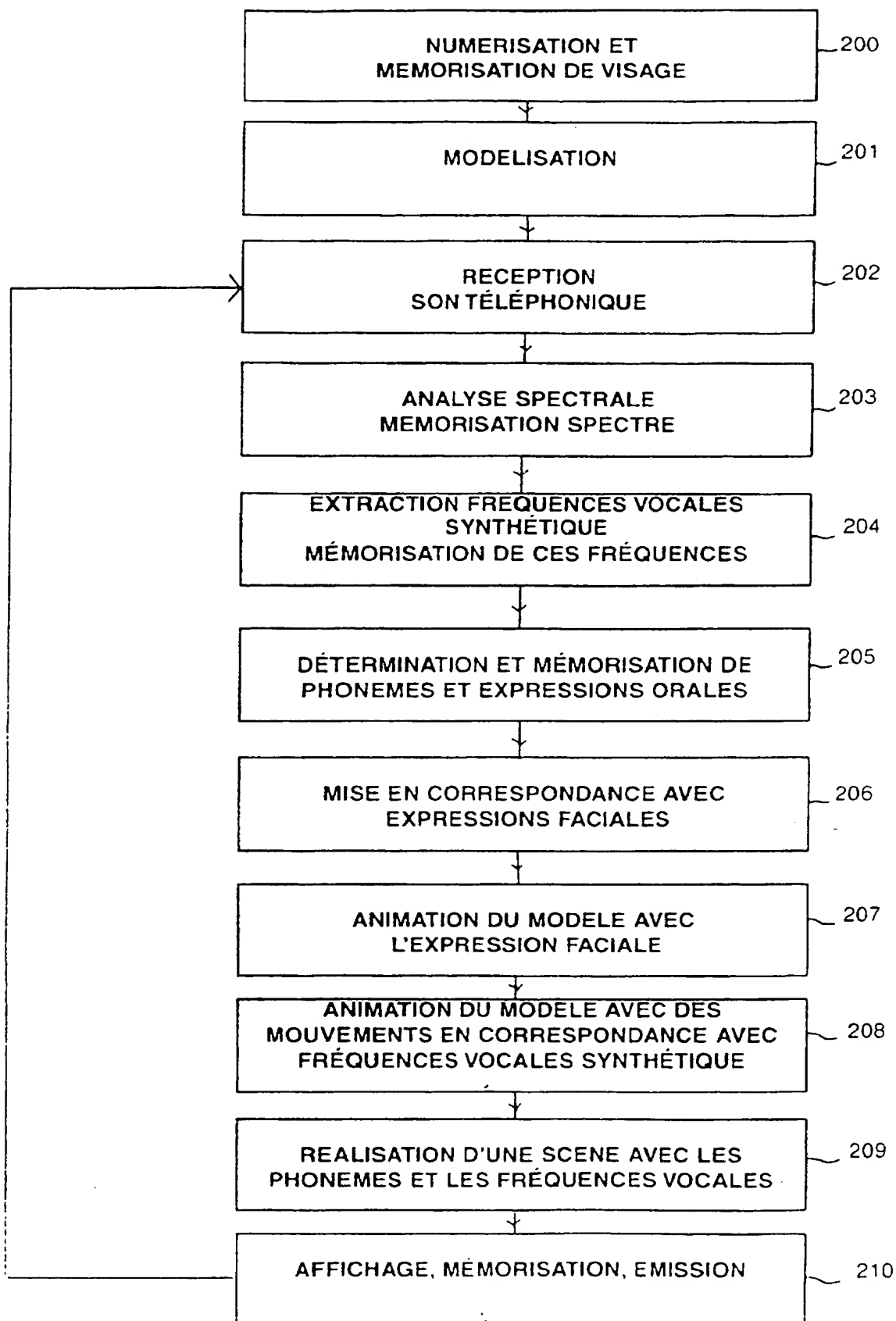


Fig. 6

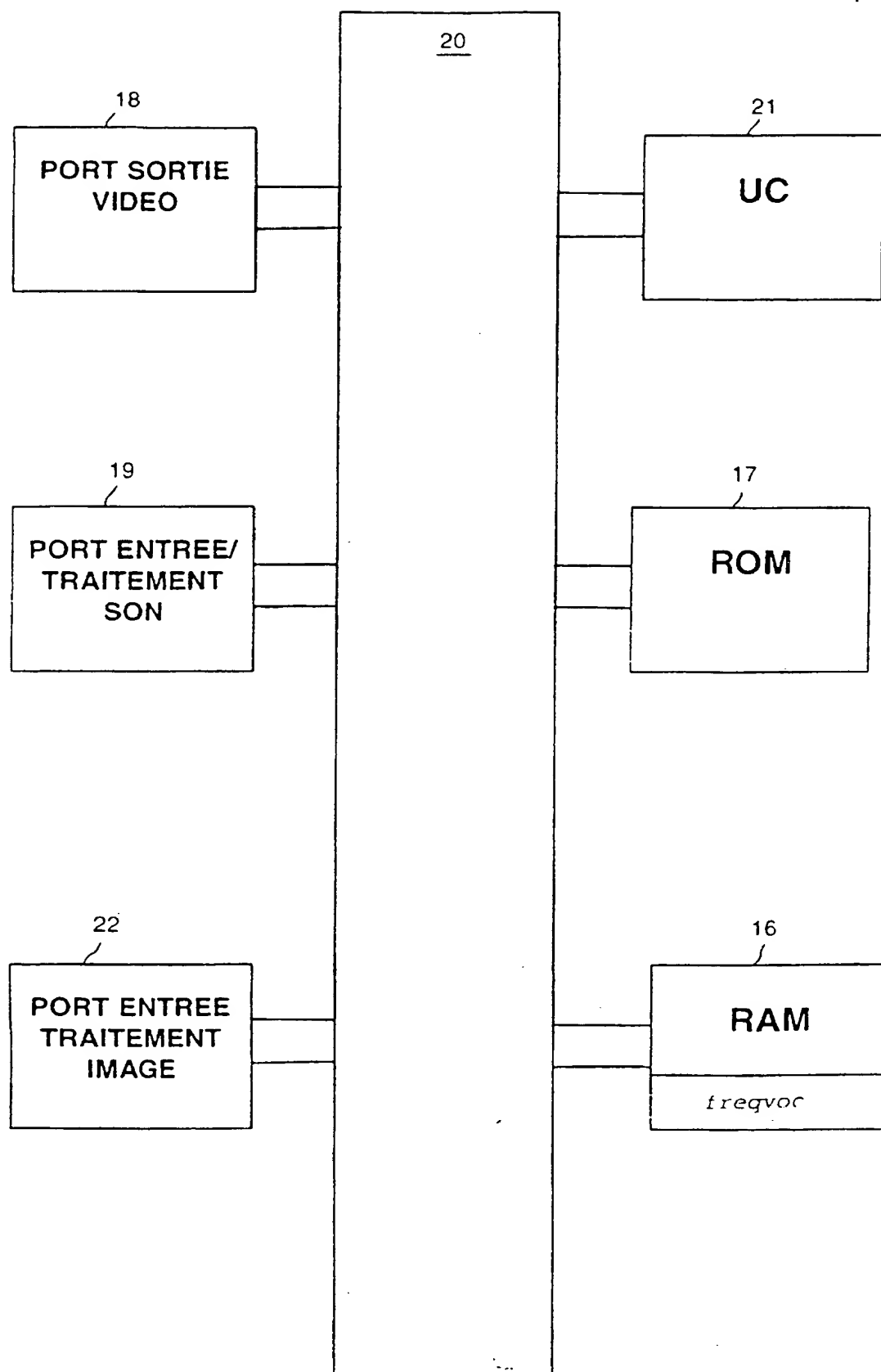


Fig. 7



Figure 8

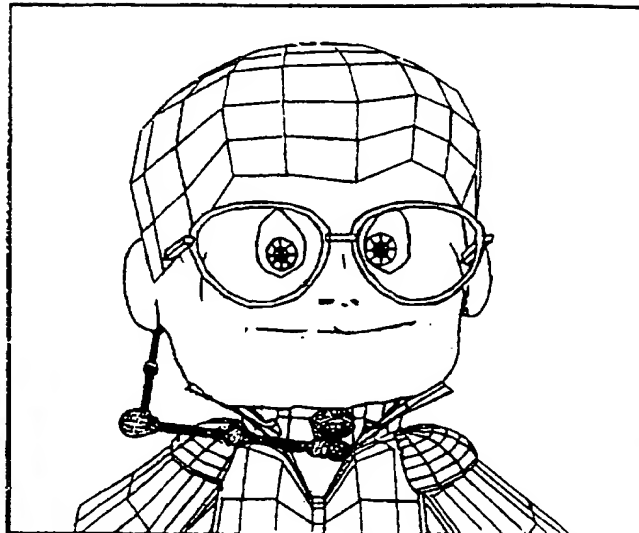


Figure 9

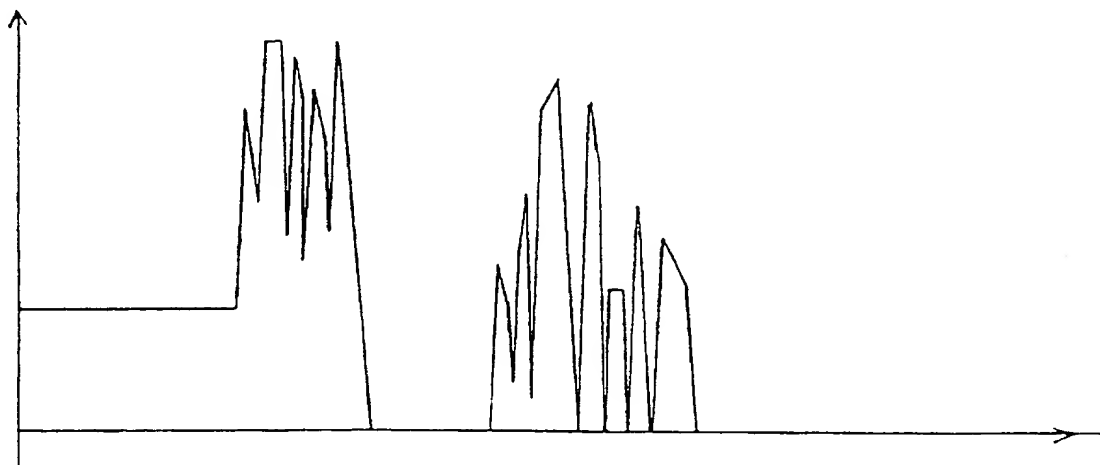


Figure 10

FEUILLE DE REMPLACEMENT (RÈGLE 26)

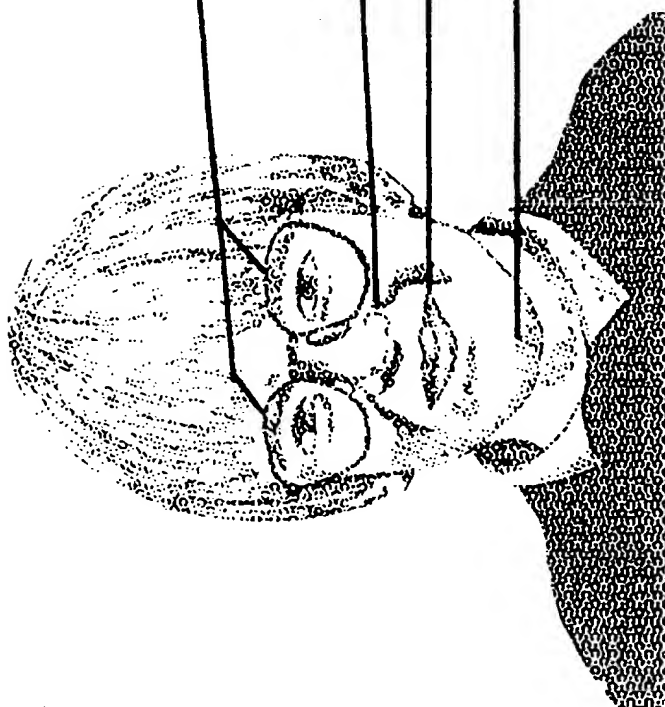


Figure 11

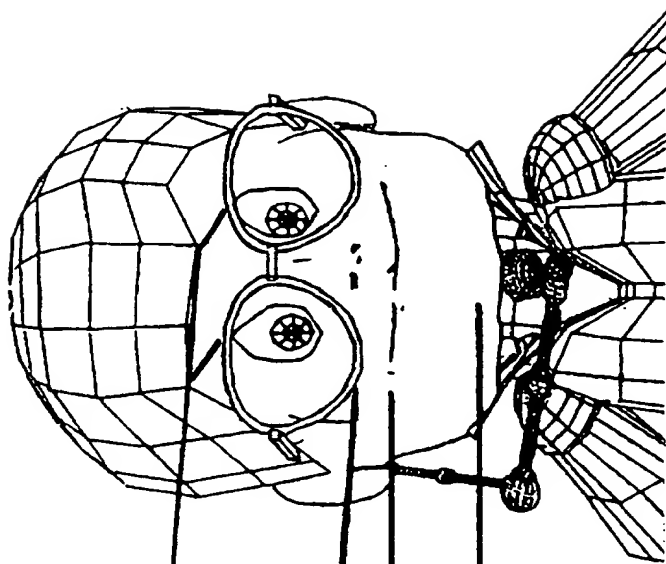


Figure 12

INTERNATIONAL SEARCH REPORT

National Application No.

PCT/FR 97/00981

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06T15/70

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>SYSTEMS & COMPUTERS IN JAPAN, vol. 22, no. 5, 1991, NEW YORK US, pages 50-59, XP000240754 MORISHIMA E.A.: "A FACIAL MOTION SYNTHESIS FOR INTELLIGENT MAN-MACHINE INTERFACE" cited in the application see page 50, left-hand column, line 1 - line 29 see page 53, left-hand column, line 11 - line 28 see page 57, right-hand column, line 1 - line 34</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	<p>1,2, 4-10,12, 15,17</p>

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

3 September 1997

Date of mailing of the international search report

23.09.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+ 31-70) 340-3016

Authorized officer

Burgaud, C

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 97/00981

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GB 2 250 405 A (BRITISH TELECOMMUNICATIONS) 3 June 1992 cited in the application see page 2, line 13 - line 29 see page 7, line 5 - line 22 ---	1-3, 5-11,13, 15,17
X	IEEE VIRTUAL REALITY ANNUAL INTERNATIONAL SYMPOSIUM, 18 - 22 September 1993, NEW-YORK NY US, pages 486-491, XP000457717 MORISHIMA E.A.: "FACIAL EXPRESSION SYNTHESIS BASED ON NATURAL VOICE FOR VIRTUAL FACE-TO-FACE COMMUNICATION WITH MACHINE" cited in the application see the whole document ---	1-3, 5-11,13, 15,17
X	INTERCHI '93 CONFERENCE PROCEEDINGS, 24 - 29 April 1993, AMSTERDAM, pages 187-193, XP000473765 TAKEUCHI AND NAGAO: "COMMUNICATIVE FACIAL DISPLAYS AS A NEW CONVERSATIONAL MODALITY" cited in the application see the whole document -----	1-3, 5-11,13, 15,17

INTERNATIONAL SEARCH REPORT
Information on patent family members

National Application No

PCT/FR 97/00981

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
GB 2250405 A	03-06-92	NONE	

Form PCT/ISA/210 (patent family annex) (July 1992)

RAPPORT RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR 97/00981

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 6 G06T15/70

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 6 G06T

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>SYSTEMS & COMPUTERS IN JAPAN, vol. 22, no. 5, 1991, NEW YORK US, pages 50-59, XP000240754 MORISHIMA E.A.: "A FACIAL MOTION SYNTHESIS FOR INTELLIGENT MAN-MACHINE INTERFACE" cité dans la demande voir page 50, colonne de gauche, ligne 1 - ligne 29 voir page 53, colonne de gauche, ligne 11 - ligne 28 voir page 57, colonne de droite, ligne 1 - ligne 34</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	<p>1,2, 4-10,12, 15,17</p>

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- * "A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- * "E" document antérieur, mais publié à la date de dépôt international ou après cette date
- * "I" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- * "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- * "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- * "T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- * "X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- * "Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- * "&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

3 Septembre 1997

Date d'expédition du présent rapport de recherche internationale

23.09.97

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+ 31-70) 340-3016

Fonctionnaire autorisé

Burgaud, C

RAPPORT DE RECHERCHE INTERNATIONALE

Deposition internationale No
PCT/FR 97/00981

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	GB 2 250 405 A (BRITISH TELECOMMUNICATIONS) 3 Juin 1992 cité dans la demande voir page 2, ligne 13 - ligne 29 voir page 7, ligne 5 - ligne 22 ---	1-3, 5-11,13, 15,17
X	IEEE VIRTUAL REALITY ANNUAL INTERNATIONAL SYMPOSIUM, 18 - 22 Septembre 1993, NEW-YORK NY US, pages 486-491, XP000457717 MORISHIMA E.A.: "FACIAL EXPRESSION SYNTHESIS BASED ON NATURAL VOICE FOR VIRTUAL FACE-TO-FACE COMMUNICATION WITH MACHINE" cité dans la demande voir le document en entier ---	1-3, 5-11,13, 15,17
X	INTERCHI '93 CONFERENCE PROCEEDINGS, 24 - 29 Avril 1993, AMSTERDAM, pages 187-193, XP000473765 TAKEUCHI AND NAGAO: "COMMUNICATIVE FACIAL DISPLAYS AS A NEW CONVERSATIONAL MODALITY" cité dans la demande voir le document en entier -----	1-3, 5-11,13, 15,17

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande Internationale No

PCT/FR 97/00981

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
GB 2250405 A	03-06-92	AUCUN	

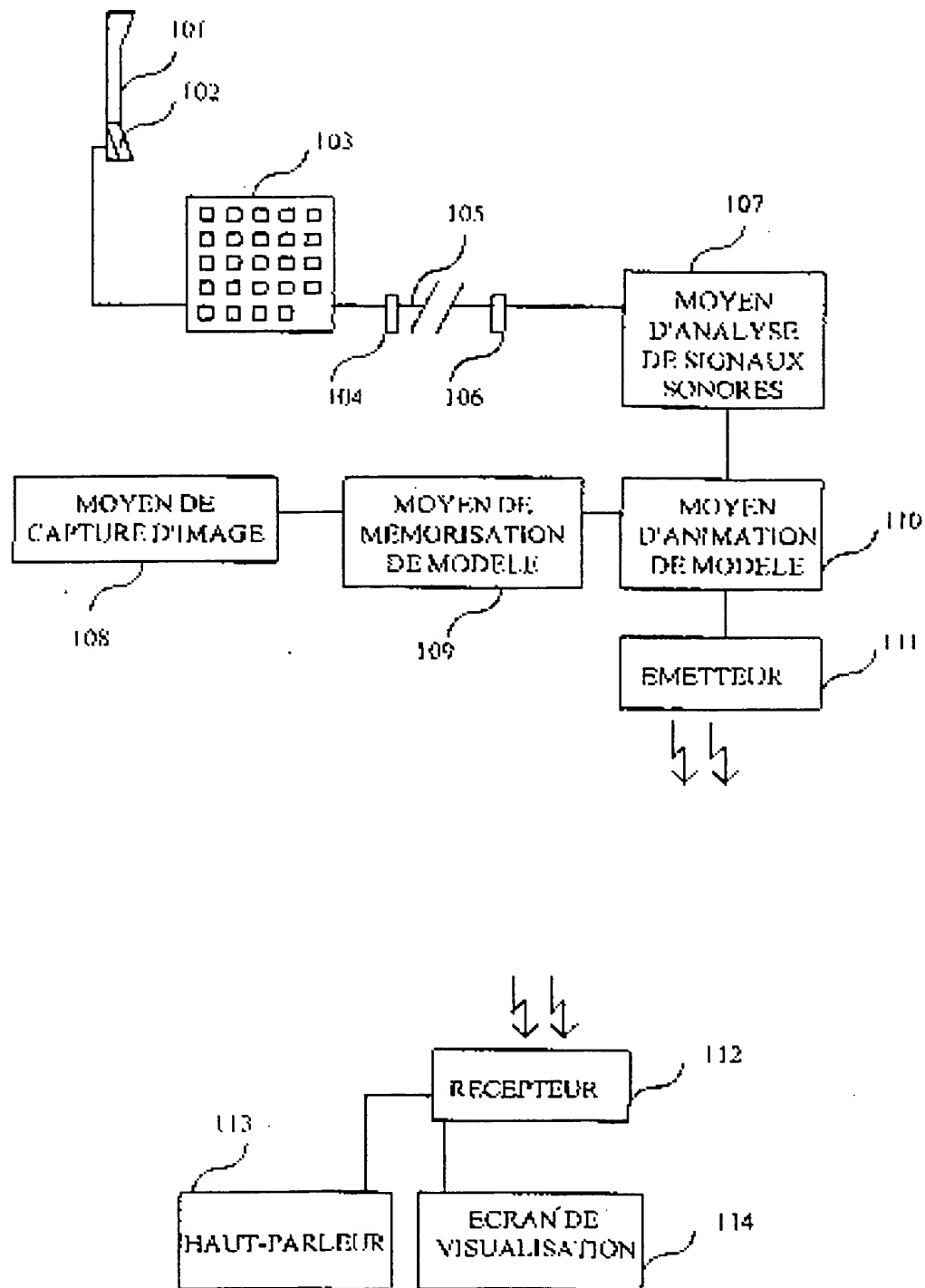


Fig. 1

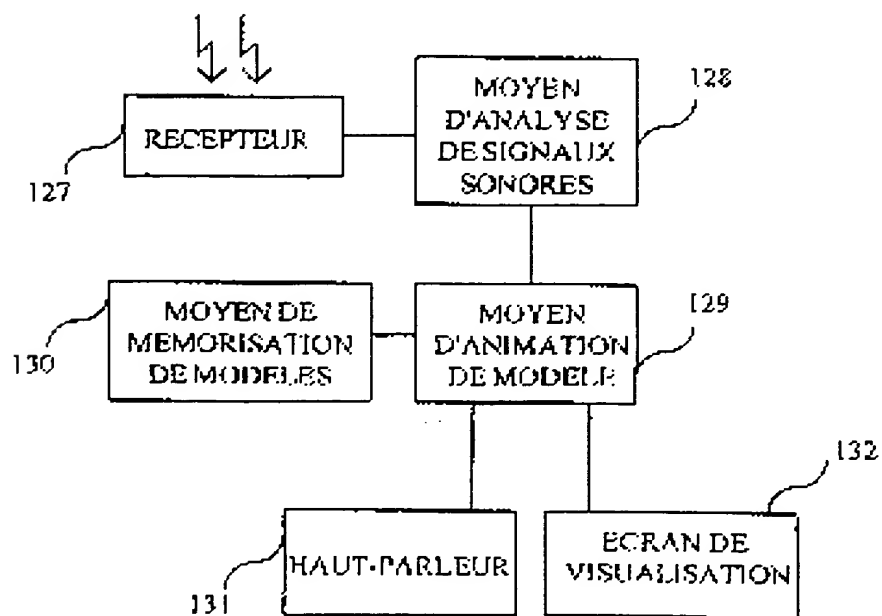
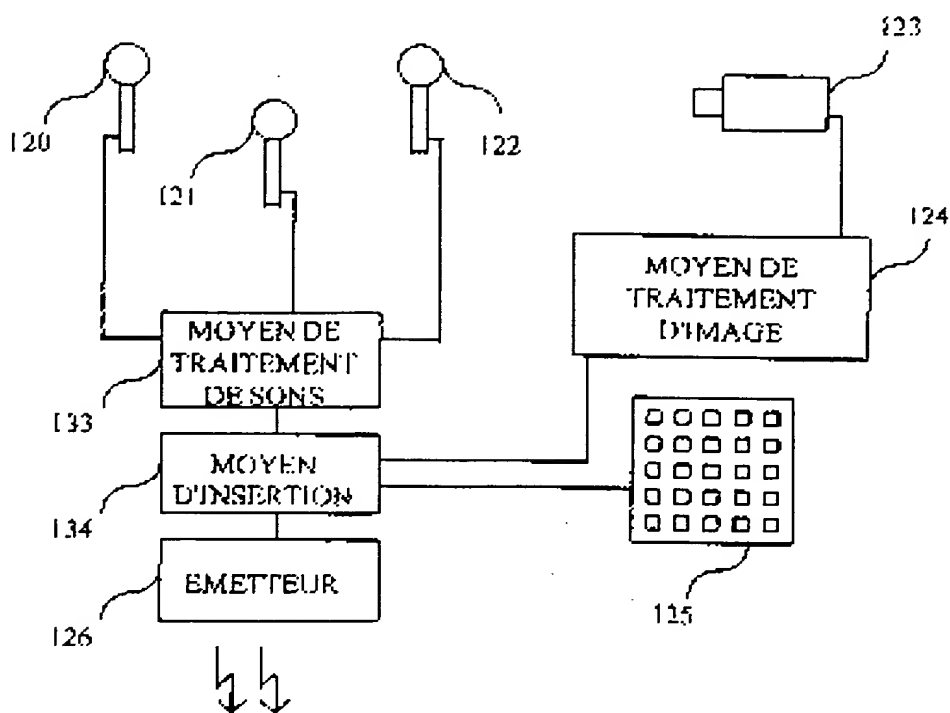


Fig. 2

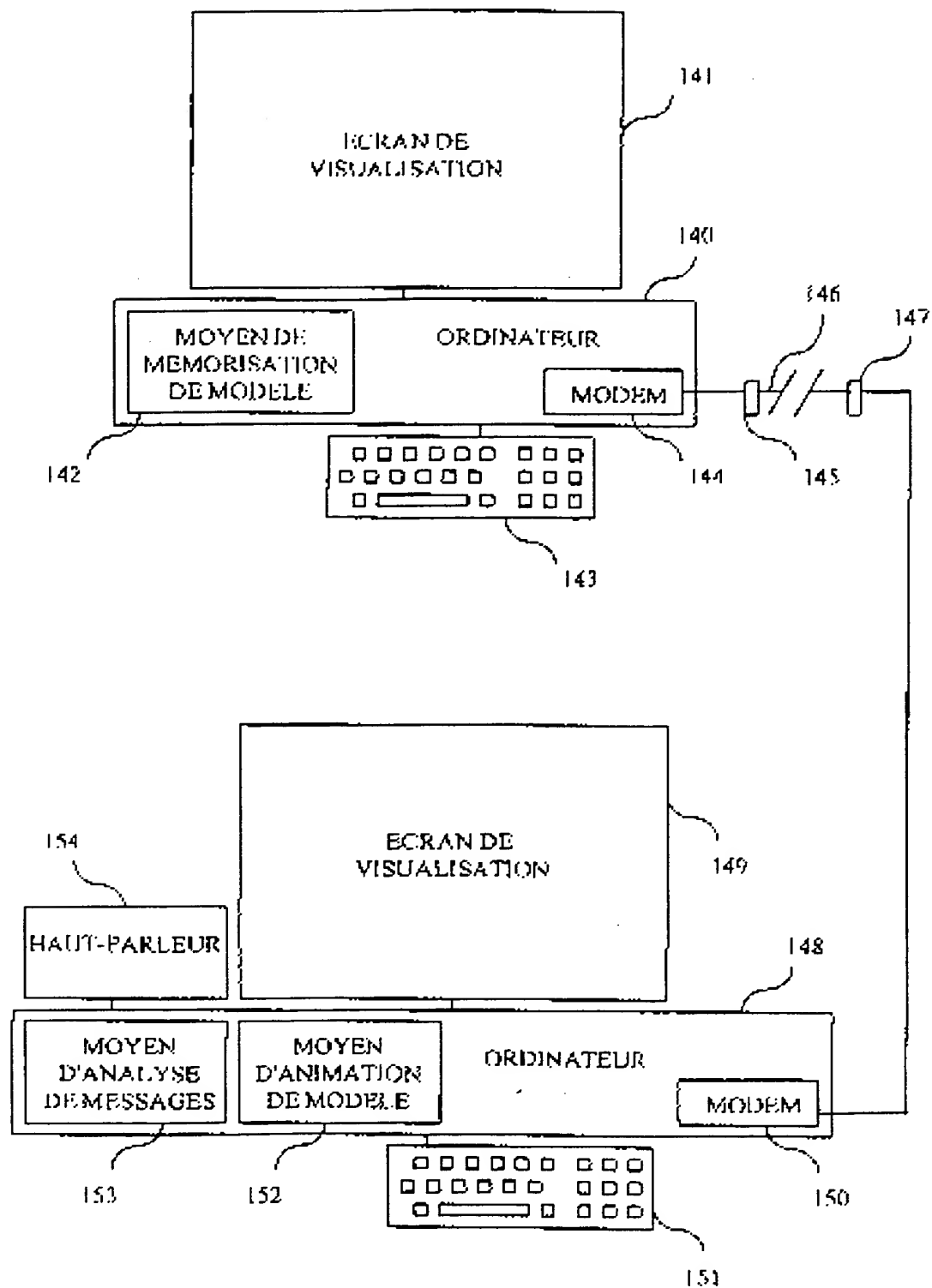


Fig. 2

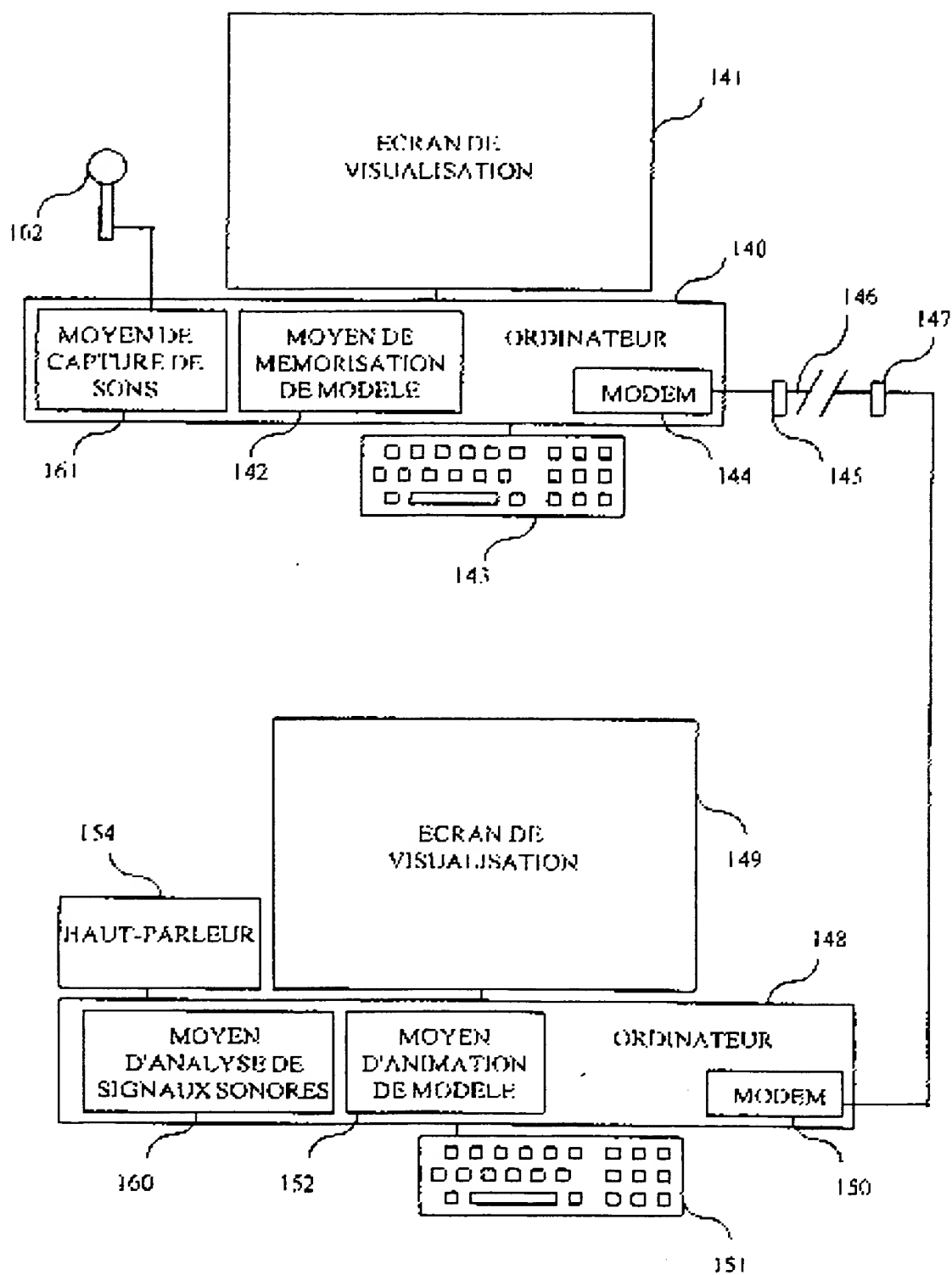


Fig. 9

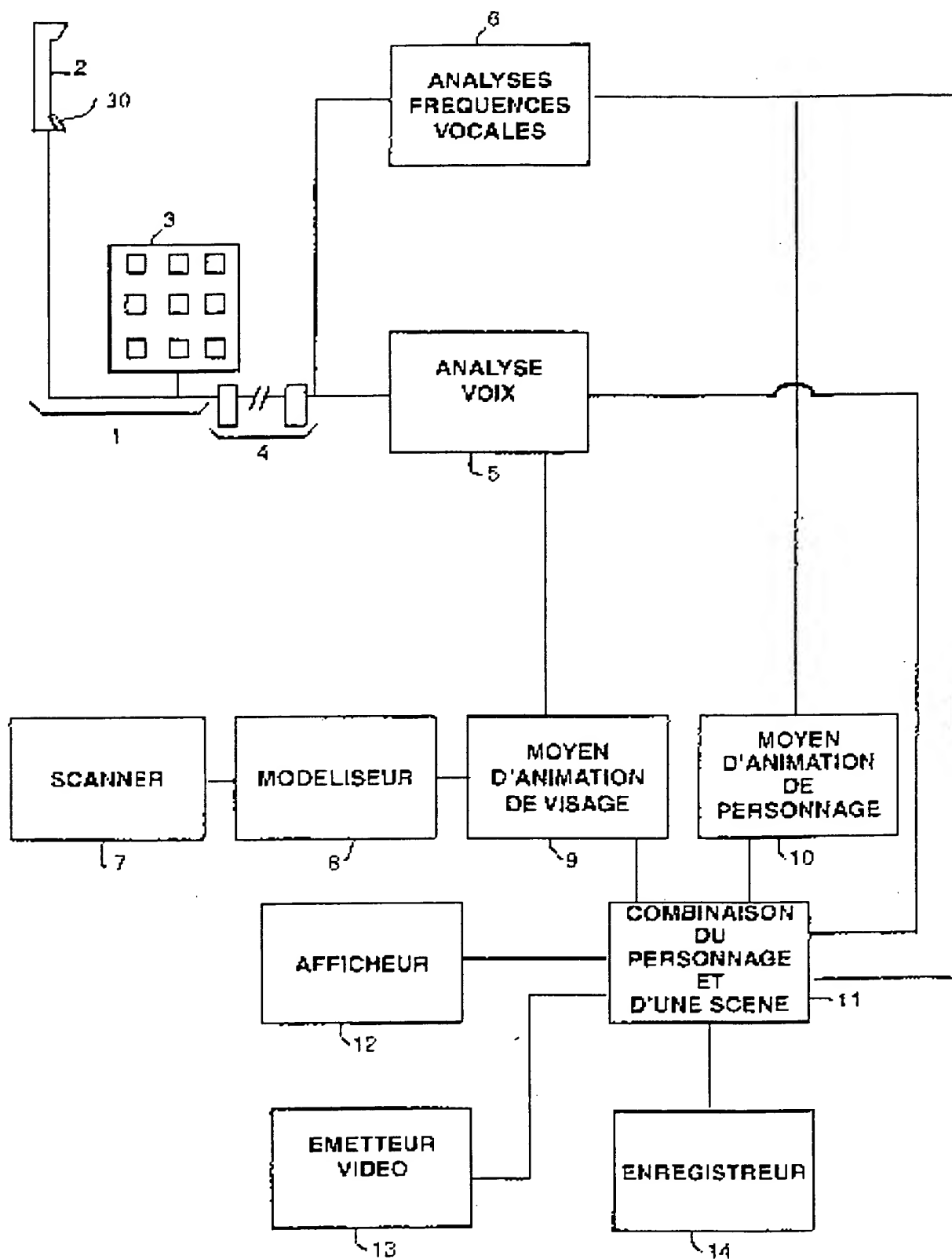


Fig. 5

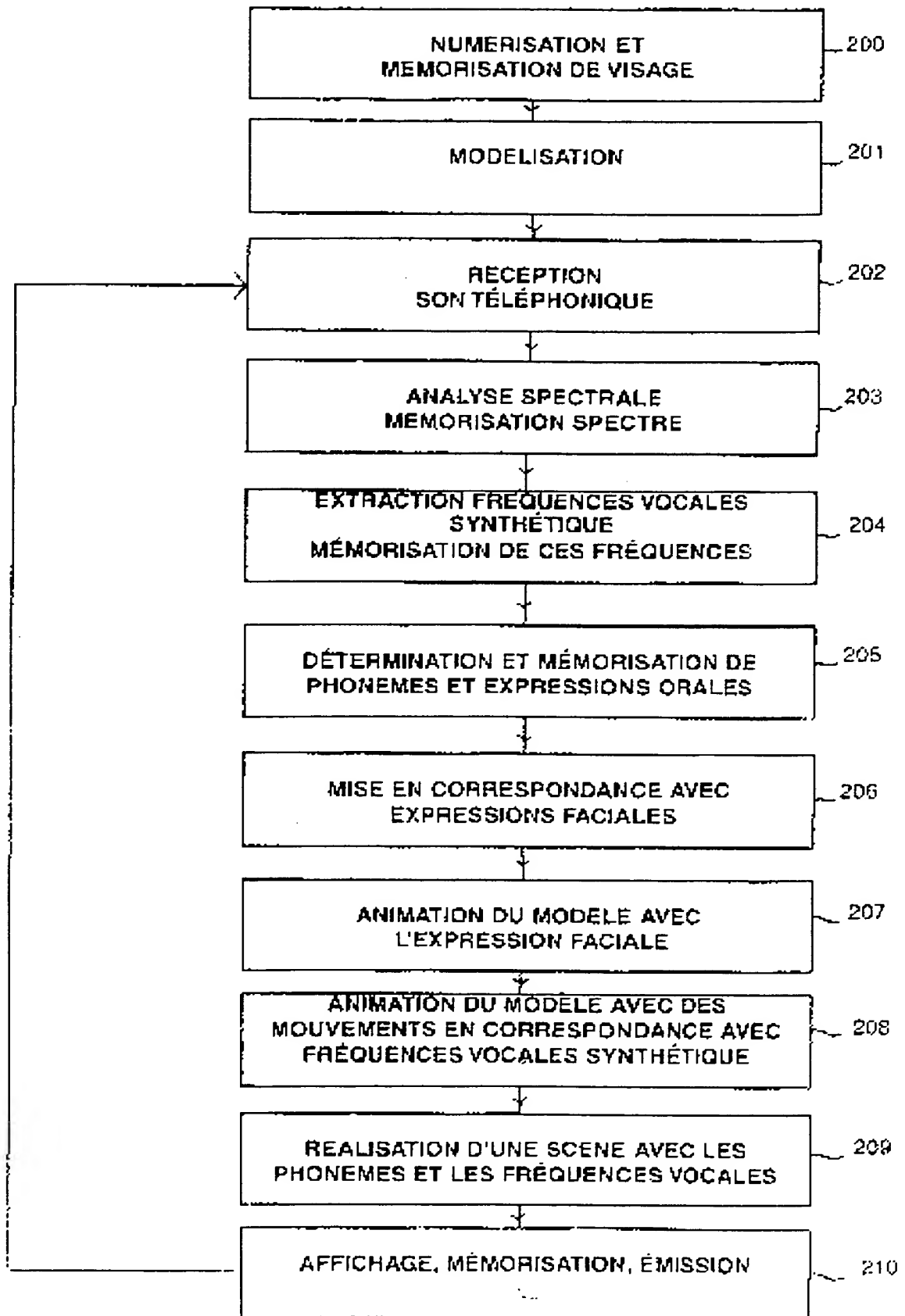


Fig. 6

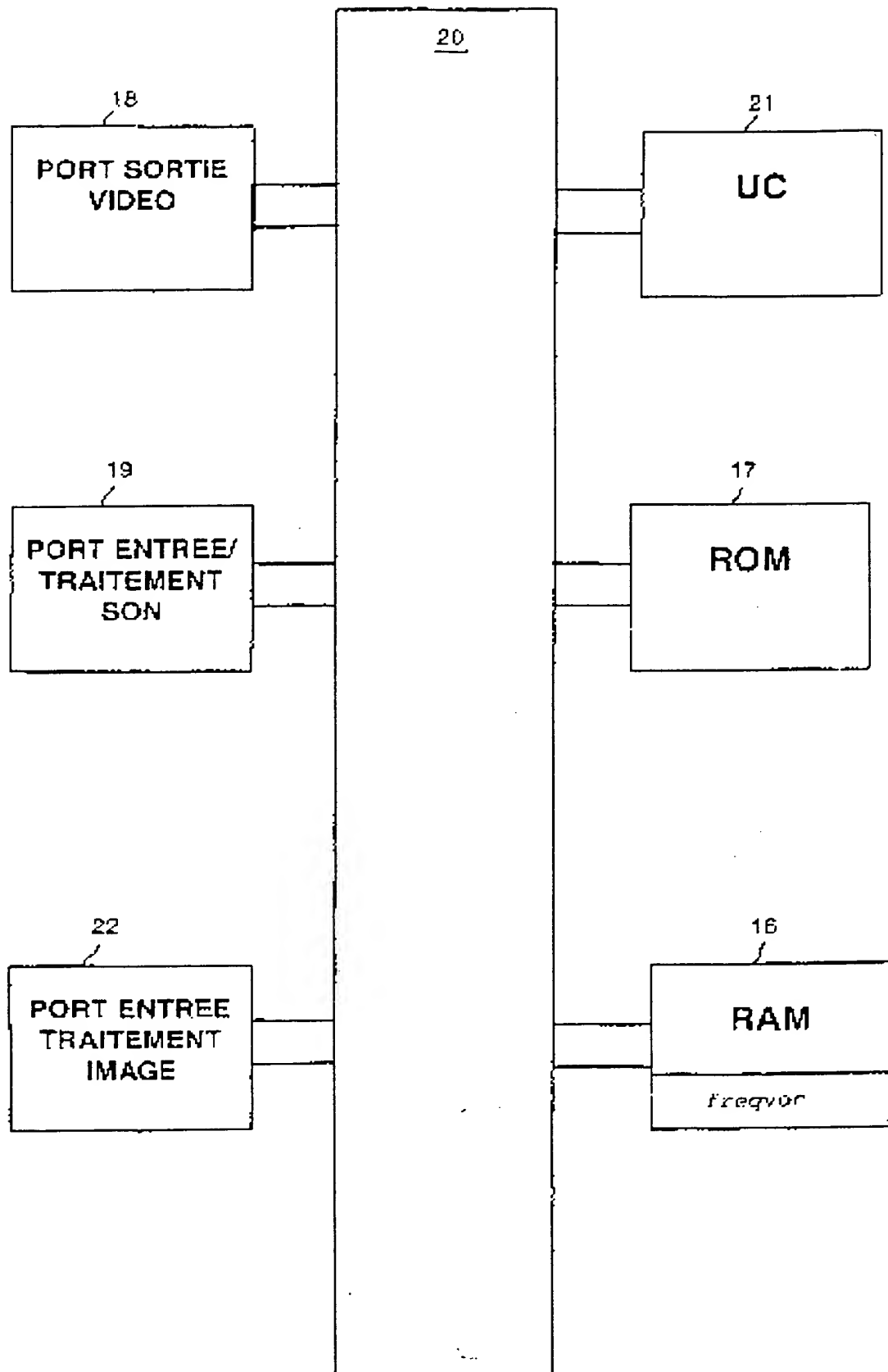


Fig. 7



Figure 8

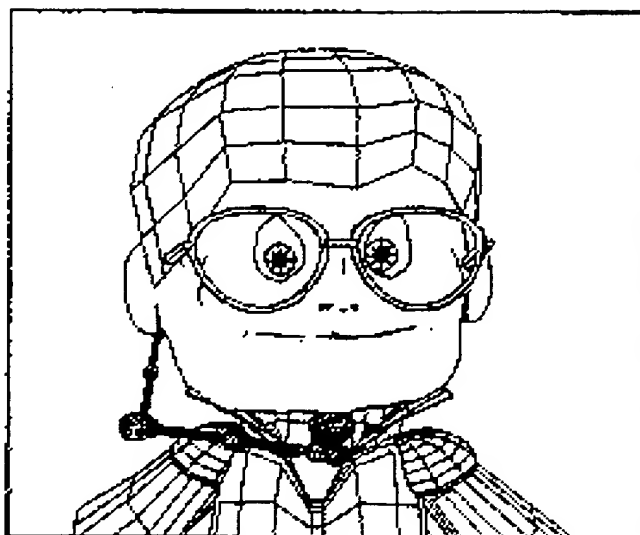


Figure 9

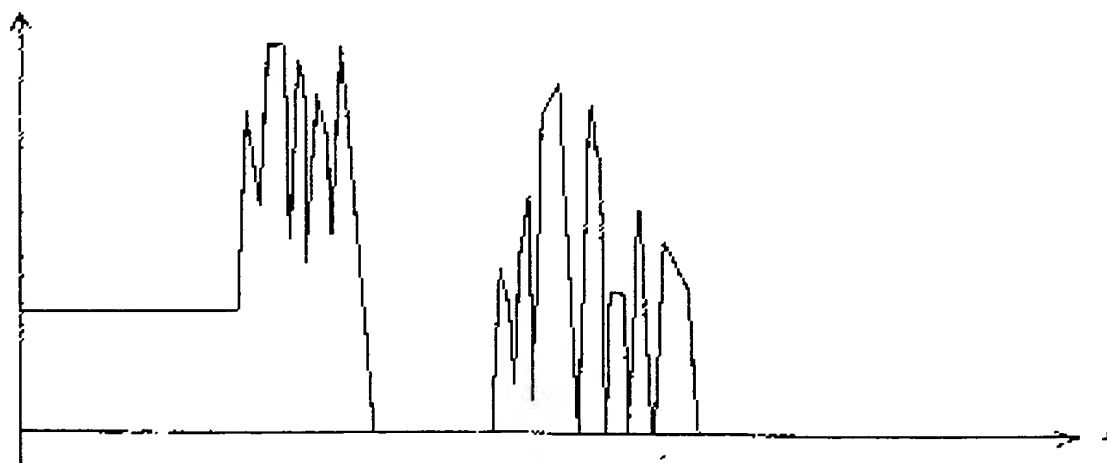


Figure 10

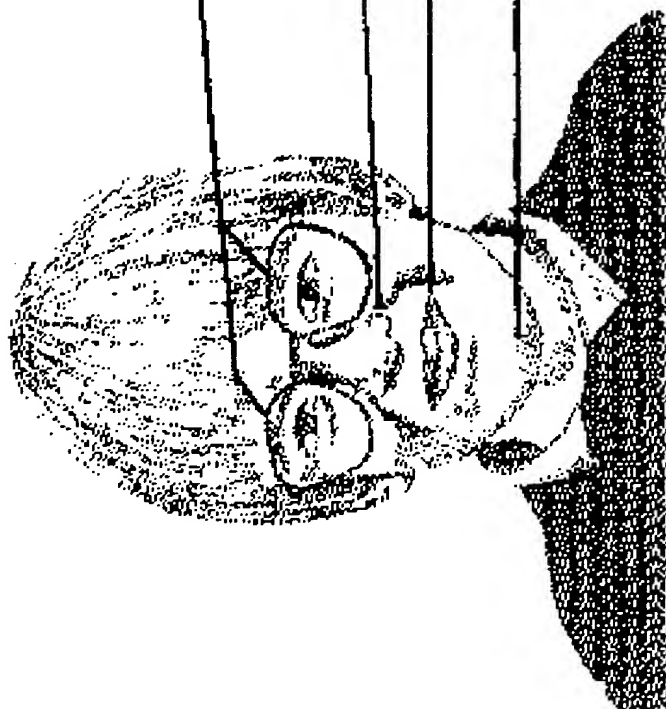


Figure 11

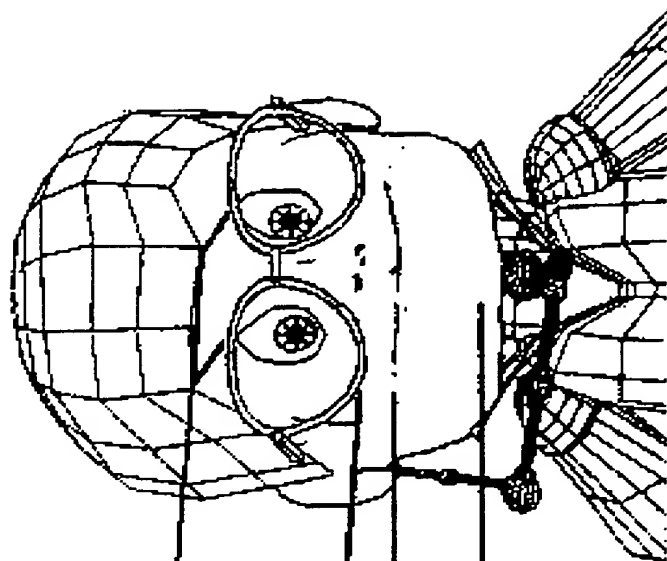


Figure 12

THIS PAGE BLANK (USPTO)